

(12) PATENT
(19) AUSTRALIAN PATENT OFFICE

(11) Application No. AU 199853272 B2
(10) Patent No. 737857

(54) Title
Non-M non-O HIV strains, fragments and applications

(51)⁷ International Patent Classification(s)
C12N 015/49 C12N 007/00
A61K 039/21 C12Q 001/68
C07K 014/16 G01N 033/50
C07K 016/10

(21) Application No: 199853272

(22) Application Date: 1997.12.08

(87) WIPO No: WO98/26075

(30) Priority Data

(31) Number	(32) Date	(33) Country
96/15087	1996.12.09	FR

(43) Publication Date : 1998.07.03

(43) Publication Journal Date : 1998.09.03

(44) Accepted Journal Date : 2001.08.30

(71) Applicant(s)
Institut National De La Sante Et De La Recherche Medicale - Inserm; Assistance
Publique - Hopitaux De Paris; Institut Pasteur

(72) Inventor(s)
Philippe Mauclere; Ibtissam Loussert-Ajaka; Francois Simon; Sentob
Saragosti; Francoise Barre-Sinoussi

(74) Agent/Attorney
GRIFFITH HACK,GPO Box 1285K,MELBOURNE VIC 3001

DE

.S (PCT)

(51) Classification internationale des brevets ⁶ : C12N 15/49, 7/00, C12Q 1/68, A61K 39/21, C07K 16/10, 14/16, G01N 33/50	A1	(11) Numéro de publication internationale: WO 98/26075 (43) Date de publication internationale: 18 juin 1998 (18.06.98)
---	----	---

(21) Numéro de la demande internationale: PCT/FR97/02227

(22) Date de dépôt international: 8 décembre 1997 (08.12.97)

(30) Données relatives à la priorité:
96/15087 9 décembre 1996 (09.12.96) FR

(71) Déposants (pour tous les Etats désignés sauf US): INSTITUT NATIONAL DE LA SANTE ET DE LA RECHERCHE MEDICALE - INSERM [FR/FR]; 101, rue de Tolbiac, F-75654 Paris Cedex 13 (FR). ASSISTANCE PUBLIQUE-HOPITAUX DE PARIS [FR/FR]; 3, avenue Victoria, F-75100 Paris RP (FR). INSTITUT PASTEUR [FR/FR]; 28, rue du Docteur Roux, F-75724 Paris Cedex 15 (FR).

(72) Inventeurs; et

(75) Inventeurs/Déposants (US seulement): MAUCLERE, Philippe [FR/FR]; 2, rue Buhan, F-33000 Bordeaux (FR). LOUSSERT-AJAKA, Ibtissam [FR/FR]; 26, avenue de la République, F-78500 Sartrouville (FR). SIMON, François [FR/FR]; 8, rue Germain Pilon, F-75018 Paris (FR). SARAGOSTI, Sentob [FR/FR]; 69 bis, rue de Billancourt, F-92100 Boulogne Billancourt (FR). BARRE-SINOUSSE, Françoise [FR/FR]; 104 Le Capricorne, 50, rue d'Erevan, F-92130 Issy-les-Moulineaux (FR).

(74) Mandataire: CABINET ORES; 6, avenue de Messine, F-75008 Paris (FR).

(81) Etats désignés: AL, AU, BB, BG, BR, CA, CN, CU, CZ, EE, GE, HU, ID, IL, IS, JP, KP, KR, LK, LR, LT, LV, MG, MK, MN, MX, NO, NZ, PL, RO, SG, SI, SK, TR, TT, UA, US, UZ, VN, brevet ARIPO (GH, KE, LS, MW, SD, SZ, UG, ZW), brevet eurasien (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), brevet européen (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), brevet OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Publiée

Avec rapport de recherche internationale.

Avant l'expiration du délai prévu pour la modification des revendications, sera republiée si de telles modifications sont reçues.

(54) Title: NON-M NON-O HIV STRAINS, FRAGMENTS AND APPLICATIONS

(54) Titre: SOUCHES DE VIH-1 NON-M NON-O, FRAGMENTS ET APPLICATIONS

(57) Abstract

The invention concerns retroviral strains of the group HIV-1, non-M non-O, particularly a strain called YBF30, its fragments and its applications as diagnosis reagent and as immunogenic agent. The HIV-2 different both from the group M and from the group O have the following characteristics: little or no serological response with respect to proteins of groups M and O and strong serological response with respect to proteins derived from the YBF30 strain or the SIV CPZGAB strain; absence of genomic amplification by the primers of regions *env* and *gag* of the HIV-1-1 of groups M and O; genomic amplification in the presence of the primers derived from the YBF30 strain; and homology of the envelope gene products higher than 70 % with respect to the YBF30 strain.

(57) Abrégé

Souches de rétrovirus du groupe VIH-1, non-M non-O, notamment une souche dénommée YBF30, ses fragments ainsi que ses applications, en tant que réactif de diagnostic et en tant qu'agent immunogène. Les VIH-1 distincts à la fois du groupe M et du groupe O présentent les caractéristiques suivantes: peu ou pas de réactivité sérologique vis-à-vis des protéines des groupes M et O et forte réactivité sérologique vis-à-vis des protéines issues de la souche YBF30 selon l'invention ou de la souche SIV CPZGAB; absence d'amplification génomique à l'aide des amorces des régions *env* et *gag* des VIH-1 des groupes M et O; amplification génomique en présence des amorces issues de la souche YBF30, selon l'invention; et homologie des produits du gène d'enveloppe supérieure à 70 % vis-à-vis de la souche YBF30.

YLO 5' AT T G C G T A C T C A C A C T T T G G
 LPB5.1 5' G G C A A G C A G G A G C T G G
 GAG Y 5' T C G T T G A G G A G T G G A C
 AS1.1 5' G G A A C A G G A G G A T T A G C A G
 GAG Y 5' G G A G C A G G C T A T G T C A C A
 GAG Y 5' G G T G A A G G C G C T A G A A G A G
 GAG Y 5' G G A C A G A G A A C T C T G T G T A G
 S1.1 5' G G A A G A A A G C A G T T G G T A G
 GAG Y 5' T T T G T T C C C T G T A T G T C
 YRT AS 1.3 5' G T T A T A T G G A T T G T G A G G
 YRT AS1.1 5' T G G A G C A C A T T A T A G T G G
 YRT 2 5' A T G A T T T A C G A G T A G T G A C G A
 YRT AS1 5' T G T C A G G G G T C G T A A G G C
 YRTB-1 5' T G G T G T G G A T G G A T A T G
 YRTB-2 5' T C T A T G A G G A G A T C A G A G
 YRT-3 5' A A T G A G A T C T G C G C A T A C
 YRTB-4 5' T G A C A G A T A G G G G A A G A C
 4401-1 5' A A C G G G A T T T G C A C T G C
 4401-2 5' A G A T G G A C G G G A G A G A G
 4201.1 5' A G C A G A C A C A T A C A G A G
 4201.2 5' A A A G T A G T C C G A C G T A G G
 4201.3 5' A T A T C C C A G T A G G T G A G G
 4201.4 5' T C T A G C A C T A A G A G G C T G
 5000.0 5' A C T C T T A G T G C T G T A G G
 5000.1 5' C G A T A G T A G A C T G T T A G C
 5000.2 5' C A T A G G T A T C G T T A C A A A G C
 5000.3 5' T C A T A A T G C A A G G C T G
 5000.4 5' G T A T G A C A T G A T G T G T C
 5000.1 5' A T T C T A G A C G A G T C C A G
 5000.1 5' C G T T A G G A T G A G A G A A T G C
 5000.2 5' T G G G A G A G T C T G T G A G G
 5000.3 5' T T T G A G C T G T G T G T G G
 LSI AS1.3 5' A T T A A G G A A G G T G A T A G C
 LSI AS1.1 5' T G T G G T T C A G G G A A G
 LSI AS1.1 5' G G T G C A T G T G A C A T A T G
 LSI A1 5' A G A G A G A C G A G T A C A G
 YLPA 5' A T A A A A G C A G G G G T G T C G

ABSTRACT

Retroviral strains of the non-M, non-O HIV-1 group, in particular a strain designated YBF30, its fragments and also its uses as a diagnostic reagent and as an immunogenic agent.

The HIV-1 viruses which differ both from the M group and the O group exhibit the following characteristics:

- * little or no serological reactivity with regard to the proteins of the M and O groups and strong serological reactivity with regard to the proteins which are derived from the strain YBF30 according to the invention or the strain CPZGAB SIV;

- * absence of genomic amplification when using primers from the env and gag regions of the M and O HIV-1 groups;

- * genomic amplification in the presence of primers which are derived from the YBF30 strain according to the invention; and

- * homology of the products of the envelope gene which is greater than 70% with regard to the YBF30 strain.



NON-M, NON-O HIV-1 STRAINS, FRAGMENTS AND USES.

The present invention relates to retroviral strains of the non-M, non-O HIV-1 group, in particular
5 a strain designated YBF30, to its fragments and to its uses as a diagnostic reagent and as an immunogenic agent.

The human acquired immunodeficiency viruses HIV-1 and HIV-2 are retrolentiviruses, which are
10 viruses found in a large number of African primates. All these viruses appear to have a common ancestor; however, it is very difficult to prejudge the period at which these different viruses became separated from this precursor. Other viruses which are more distant,
15 but which nevertheless belong to the same group, are found in other mammals (ungulates and felines).

All these viruses are associated with long infections; an absence of symptoms is the rule in monkeys which are infected naturally.

20 While the origin of HIV-2 appears to be clear on account of its strong homology with the Sooty Mangabey (West Africa) virus, no virus which is closely related to HIV-1 has been found in monkeys. The most closely related viruses are viruses found in two
25 chimpanzees (CPZGAB SIV, ANT SIV).

All the lentiviruses have been found to exhibit substantial genetic variability, and the phylogenetic study of these variants, obtained from a large number of different geographic locations, has enabled 8
30 subtypes (clades) of HIV-1 to be distinguished, all of which are equidistant from each other. The clades are only a mathematical representation of the expression of the variability: phenetic analysis, which is based on the amino acids rather than on the nucleic acids, gives
35 different results (Korber et al., 1994).

The demonstration of subtypes is in accord with a phylogenetic analysis which does not, to date, have any pathophysiological correlation but, instead, a geographical correspondance. This is because each



subtype is mainly found in a particular geographical area. The B subtype is predominant in Europe and the United States whereas two subtypes, i.e. E and B, are found in Thailand and there is a strong correlation
5 between the mode of transmission which, in actual fact, corresponds to a particular population and the subtype found. All the clades have been found in Africa and their distribution across the rest of the world reflects a probability of encounter between persons
10 indulging in high-risk behaviour. The main clade, which is the main one because it is present in substantial proportions in Africa, is clade A. A very great degree of variability has been found in some African countries (G. Myers, 1994; P.M. Sharp et al., 1994). Several
15 subtypes have been characterized in the western central African countries such as the Central African Republic (Murphy et al., 1993) and Cameroon (Nkengasong et al., 1994).

Finally, patients have been characterized who
20 are carriers of viral variants of HIV-1, whose sera have posed detection problems for particular kits which are sold on the French market and whose confirmatory Western blots have been atypical (Loussert-Ajaka et al., 1994; Simon et al., 1994; PCT International
25 Application WO 96/27013).

Analysis of these variants has confirmed the fact that the type 1 HIV viruses should be subdivided into two groups, i.e. the M (major) group and an O (outlier) group, which includes these isolates, as
30 Charneau et al., 1994 had proposed. Analysis of the synonymous mutations/non-synonymous mutations ratio carried out on the sequences of the known O group viruses indicates that this new group is also ancient, even if no more ancient than the M group (Loussert-Ajaka et al., 1995). Its low prevalence to date, i.e.
35 8% of patients infected with HIV-1 in Cameroon (Zekeng et al., 1994) and 18 cases characterized in France, is thought to be due to factors which are purely epidemiological.



These two groups of HIV-1 form a tree which is in the shape of a double star (Figures 9 to 19). Two isolates, i.e. CPZGAB SIV, characterized from a chimpanzee from Gabon (Huet et al., 1990) and CPZANT SIV, characterized from a chimpanzee in the Antwerp Zoo, possess sequences and genetic organizations which are very closely related to HIV-1 but which do not fall within either of these two groups and form two new branches on the phylogenetic tree.

10 The demonstration of new variants is important for developing sufficiently sensitive and specific reagents for detecting HIV infections, that is to say reagents which do not lead to false-negative or false-positive results, and for developing compositions which
15 are protective in regard to subtypes which do not belong either to the M group or to the O group.

Consequently, the invention provides a non-M, non-O strain, as well as sequences derived from this strain, which are suitable for detecting non-M and non-O HIV-1
20 variants and which do not lead to false-negative or false-positive results being obtained. In order to do this, the inventors have, in particular, established an algorithm for differentiating between, and confirming,
25 group M and group O HIV-1 infections, thereby enabling them to select non-M, non-O variants.

The present invention relates to a non-M, non-O HIV-1 strain which exhibits the morphological and immunological characteristics of the retrovirus which
30 was deposited on 2 July 1996 under number I-1753 (designated YBF30) in the Collection Nationale de Cultures de Microorganismes (National Collection of Microorganism Cultures), kept by the Pasteur Institute.

A non-M, non-O variant is understood as meaning
35 a type 1 HIV which cannot serologically and molecularly be recognized as belonging to either of these groups.

The present invention also relates to the complete nucleotide sequence of the strain as defined above (SEQ ID No. 1) as well as to nucleic acid



fragments which are at least 10 nucleotides in size and which are derived from the said strain.

Fragments of this type which may be mentioned are:

- 5 - YBF 30 LTR (SEQ ID No. 2),
 - YBF 30 GAG (SEQ ID No. 3) (gag gene),
 - YBF 30 POL (SEQ ID No. 5) (pol gene),
 - YBF 30 VIF (SEQ ID No. 7) (vif gene),
 - YBF 30 VPR (SEQ ID No. 9) (vpr gene),
10 - YBF 30 VPU (SEQ ID No. 11) (vpu gene),
 - YBF 30 TAT (SEQ ID No. 13) (tat gene),
 - YBF 30 REV (SEQ ID No. 15) (rev gene),
 - YBF 30 ENV gp160 (SEQ ID No. 17) (env gene),
 - YBF 30 NEF (SEQ ID No. 19) (nef gene),
15 - the SEQ ID Nos. 21-57, also designated,
 respectively, YLG, LPBS.1, GAG Y AS1.1, GAG Y AS1, GAG
 6, GAG Y S1, GAG Y S1.1, GAG Y S1.2, YRT AS1.3, YRT
 AS1.2, YRT AS1.1, YRT 2, YRT AS1, YRT 2.1, YRT 2.2, YRT
20 2.3, YRT 2.4, 4481-1, 4481-2, 4235.1, 4235.2, 4235.3,
 4235.4, SK69.6, SK69.5, SK69.4, SK69.3, SK69.2, SK69.1,
 SK68.1, SK68.2, SK68.3, LSI AS1.3, LSI AS1.2, LSI
 AS1.1, LSI A1, YLPA.

25

Such sequences can be used in the specific identification of a non-M, non-O HIV-1, and as diagnostic reagents, either alone or pooled with other reagents, for the differential identification of any HIV-1.

These sequences may, in particular, be employed in diagnostic tests which comprise either a direct hybridization with the viral sequence to be detected or an amplification of the said viral sequence, with these tests using, as primers or as probes, an oligonucleotide which comprises at least 10 nucleotides and which is included in any one of the above



sequences, in particular one of the abovementioned sequences, SEQ ID Nos. 21-57.

The present invention also relates to HIV-1 viruses which are characterized in that they differ
5 both from the M group and from the O group and exhibit the following characteristics:

• little or no serological reactivity with regard to proteins of the M and O groups and strong serological reactivity with regard to proteins which
10 are derived from the YBF30 strain or the CPZGAB SIV strain;

• absence of genomic amplification when using primers from the env and gag regions of HIV-1 viruses of the M and O groups;

15 • genomic amplification in the presence of primers which are derived from the YBF30 strain, as defined above; and

• homology of the products of the envelope gene which is > 70% with regard to the YBF30 strain.

20 The invention also relates to the use of the above described sequences for implementing a method of hybridization and/or of gene amplification of nucleic acid sequences of the HIV-1 type, with these methods being applicable to the in-vitro diagnosis of the
25 potential infection of an individual with a virus of the non-M, non-O HIV-1 type.

This in-vitro diagnostic method is carried out using a biological sample (serum or circulating lymphocyte) and comprises:

30 • a step of extracting the nucleic acid which is to be detected and which belongs to the genome of the virus, which virus may possibly be present in the biological sample, and, where appropriate, a step of treating the nucleic acid using a reverse
35 transcriptase, if this nucleic acid is in RNA form,

• at least one cycle comprising the steps of denaturing the nucleic acid, of hybridizing with at least one sequence in accordance with the invention and, where appropriate, extending the hybrid, which has



been formed, in the presence of suitable reagents (polymerizing agent, such as DNA polymerase and dNTP), and

5 . a step of detecting the possible presence of the nucleic acid belonging to the genome of a virus of the non-M, non-O HIV-1 group type.

The following conditions are employed for the PCR using the primers derived from the YBF30 strain:

10 - extracting the lymphocytic DNA by means of the phenol/chloroform technique and quantifying it by spectrophotometry at a wavelength of 260 nm. All the amplifications are carried out using a Perkin Elmer 2400 thermocycler.

15 - the long (9 kb) PCRs are carried out using an XL PCR kit (Perkin Elmer) in accordance with the manufacturer's conditions and using the dNTP's, the buffers provided and Perkin Elmer's "hot start"; the amplification cycles of this long PCR are:

20 . 1 cycle of denaturation for 2 minutes at 94°C,

. then 16 cycles: 15 seconds at 94°C, 15 seconds at 55°C, 8 minutes at 68°C,

25 . then 24 cycles: 15 seconds at 94°C, 15 seconds at 55°C, 8 minutes at 68°C, adding a further 15 seconds (incrementation) to each cycle.

- the nested PCRs are carried out on the amplification products of the long PCRs. The conditions for carrying out the nested PCRs are as follows:

30 . "Expand High Fidelity PCR System" Taq polymerase buffer and enzyme from Boehringer Mannheim in accordance with the manufacturer's instructions, dNTP and "hot start" from Perkin Elmer,

35 . 200 µmol of each dNTP, 20 pmol of each primer in accordance with the invention, 5 µl of DNA, 10 µl of 10 × PCR buffer and 2.6 units of Taq polymerase in a volume of 100 µl,

. amplification: one cycle of 2 minutes at 94°C followed by 38 cycles: 15 seconds at 94°C, 15 seconds at 55°C, a time of elongation at 72°C which varies in



accordance with the size of the PCR product to be amplified (from 30 seconds to 2 minutes) and a final elongation cycle of 10 minutes at 72°C.

5 The amplified product is preferably detected by direct sequencing.

The invention also relates to a peptide or a peptide fragment which is characterized in that it can be expressed by a non-M, non-O HIV-1 strain or using a nucleotide sequence as defined above, and in that it is
10 capable: (1) of being recognized by antibodies which are induced by a non-M, non-O HIV-1 virus, as defined above, in particular the YBF30 strain or a variant of this strain, and which are present in a biological sample which is obtained following an infection with a
15 non-M, non-O HIV-1 strain, and/or (2) of inducing the production of anti-non-M, non-O HIV-1 antibodies.

Peptides of this type which may be mentioned are, in particular, those which are derived from the YBF30 strain, in particular: that which is expressed by
20 the *gag* gene (SEQ ID No. 4), that which is expressed by the *pol* gene (SEQ ID No. 6), that which is expressed by the *vif* gene (SEQ ID No. 8), that which is expressed by the *vpr* gene (SEQ ID No. 10), that which is expressed by the *vpu* gene (SEQ ID No. 12), that which is
25 expressed by the *tat* gene (SEQ ID No. 14), that which is expressed by the *rev* gene (SEQ ID No. 16), that which is expressed by the *env* gene (SEQ ID No. 18), or one of its fragments such as a fragment of the V3 loop region, i.e. CTRPGNNTGGQVQIGPAMTFYNIKIVGDIRQAYC (SEQ
30 ID No. 58), and that which is expressed by the *nef* gene (SEQ ID No. 20), or a fragment of these peptides which are capable of recognizing the antibodies which are produced during an infection with a non-M, non-O HIV-1 as defined above.

35 The invention also relates to immunogenic compositions which comprise one or more translation products of the nucleotide sequences according to the invention and/or one of the peptides as defined above, obtained, in particular, by synthetic means.



The invention also relates to the antibodies which are directed against one or more of the above-described peptides and to their use for implementing methods for the in-vitro, in particular differential, diagnosis of the infection of an individual with a virus of the HIV-1 type using methods which are known to the skilled person.

The present invention encompasses all the peptides which are capable of being recognized by antibodies which are isolated from an infectious serum which is obtained after an infection with a non-M, non-O HIV-1 strain, and the peptides which are capable of being recognized by an antibody according to the invention.

The invention furthermore relates to a method for the in-vitro diagnosis of a non-M, non-O HIV-1 virus, which method is characterized in that it comprises bringing a biological sample, which has been taken from a patient, into contact with antibodies according to Claim 10, which may possibly be combined with anti-CPZGAB SIV antibodies, and detecting the immunological complexes which are formed between the HIV-1 antigens, which may possibly be present in the biological sample, and the said antibodies.

The invention also relates to a kit for diagnosing HIV-1, which kit is characterized in that it includes at least one reagent according to the invention.

Apart from the provisions which have been described above, the invention also comprises other provisions which will be evident from the description which follows and which refers to examples of implementing the method which is the subject of the present invention and also to the attached drawings, in which:

- Figures 1 to 7 illustrate the location of the different primers on the genome of the YBF30 strain;

- Figure 8 illustrates the genomic organization of the YBF30 strain;



- Figures 9 to 16 depict the phylogenetic analysis of the different genes of the YBF30 strain as compared with group M HIV-1 and group O HIV-1 (Figure 9: *ltr* gene, Figure 10: *gag* gene, Figure 11: *tat* gene, Figure 12: *rev* gene, Figure 13: *vif* gene, Figure 14: *env gp120* gene, Figure 15 *env gp41* gene, Figure 16: *nef* gene, Figure 17: *pol* gene, Figure 18: *vpr* gene, Figure 19: *vpu* gene); - Figure 20 illustrates the percentage genetic distance between YBF30 and HIV-1/CPZGAB SIV.

10 It should of course be understood, however, that these examples are give solely by way of illustrating the subject-matter of the invention, of which they in no way constitute a limitation.

15 For the purposes of this specification it will be clearly understood that the word "comprising" means "including but not limited to", and that the word "comprises" has a corresponding meaning.

20 All references, including any patents or patent applications, cited in this specification are hereby incorporated by reference. No admission is made that any reference constitutes prior art. The discussion of the references states what their authors assert, and the applicants reserve the right to challenge the accuracy and pertinency of the cited documents. It will be clearly understood that, although a number of prior art publications are referred to herein, this reference does not constitute an admission that any of these documents forms part of the common general knowledge in the art, in Australia or in any other country.

30 **EXAMPLE : Obtaining a non-M, non-O HIV-1 variant according to the invention (YBF30) and its uses.**

35 This was, in particular, possible in connection with studying the epidemiology of infection with human acquired immunodeficiency viruses (HIV) in Cameroon, which epidemiology is especially paradoxical. In this country,

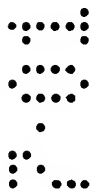


the diversity of the strains is remarkable as most of the subtypes of the M (major) group of HIV-1 viruses known to date have been reported. Cases of infection with highly divergent HIV-1 viruses of the O group (O for outlier)

- 5 have been reported, almost exclusively in patients of Cameroonian origin. Cases of infection with HIV-2, HTLV-1 and HTLV-2 subtypes A and B have also been reported.

- Taking a basis the results of previous serological and genotypic assessments, the inventors
10 established an algorithm for differentiating between and confirming infections with HIV-1 viruses of the M and O groups in order to select non-M, non-O variants.

- These methods were applied to samples which were went to the National Reference Laboratory for HIV
15 infections at Yaoundé and made it possible to characterize a highly divergent HIV isolate and to define the tools for characterizing a new HIV-1 group,



taking into account the homologies which were observed between this human strain YBF30 and the simian strain CPZGAB SIV.

5 I- Way of serologically characterizing the YBF30 variant during the epidemiological study.

1) Collecting the samples:

10 All the adult patient sera which were sent to the Yaoundé reference laboratory in 1994 and 1995 for detecting or confirming an HIV infection were studied (n = 8831).

2) Differentiating serologically between group M and group O HIV-1, and selecting variants:

15 If there was positive detection of anti-HIV antibodies (Généclavia Mixt indirect mixed HIV-1 and HIV-2 EIA, Sanofi-Pasteur, Paris, France), this was then combined with an EIA test based on the principle of competition with a specific antigen of the M group (Wellcozyme Rec HIV-1, Murex, Dartford, UK).

20 If the competitive Wellcozyme Rec HIV-1 test is positive, with a ratio for the reactivity in optical density (OD) as compared with the threshold or cut-off (CO) value which is greater than 5 ($CO/OD > 5$), the serum is regarded as being HIV-1-positive, a result which should be confirmed on a new sample.

25 The choice of a reactivity ratio which is greater than 5 for regarding the competitive test as being a test for confirming infection with HIV-1 is based on experience acquired by the virology laboratory of Bichat hospital: all of 7200 samples which reacted with a ratio > 5 gave a strongly positive HIV-1 Western blot (WB, New Lav Blot 1, SDP, Marnes la Coquette).
30 Apart from cases of HIV-1 seroconversion, the samples which are confirmed as being HIV-positive and which give a Wellcozyme ratio of < 5 correspond either to
35 infections with HIV-2 or to infections with O group HIV-1 or other HIV-1 variants.

In order to eliminate the false-positive reactions when carrying out a mixed EIA detection, the samples which give a CO/OD ratio of < 5 are tested



systematically with a third generation mixed HIV-1/HIV-2 EIA (Enzygnost Plus, Marburg, Germany) which includes antigens of the M and O HIV-1 groups (recombinant gp41 of the MVP5180 strain). If this test is positive, a rapid test which discriminates between HIV-1 and HIV-2 (Multispot, SDP, Marnes la Coquette) and a Western blot (WB, New Lav Blot 1 or 2, SDP) are then carried out.

3) Serologically confirming infections with O group HIV-1 and HIV-1 variants.

All the samples which give a CO/OD ratio of < 5, and which have been differentiated as being positive by WB (positivity criteria: 2 ENV +/- POL +/- GAG or 1 ENV + POL +/- GAG) and HIV-1, are tested with a dot blot test using peptide antigens of the V3 and transmembrane regions (InnoLia, Innogenetics, Ghent, Belgium).

4) Retroviral isolation of the group O and variant strains.

The peripheral blood mononuclear cells (PBMC) from the seropositive patients were isolated by Ficoll-Hypaque gradient in Cameroon and then stored, and transported to Paris, in liquid nitrogen.

After thawing, the PBMCs from the patients were cocultured together with lymphocytes from seronegative Caucasian donors. Viral replication in the culture supernatants was demonstrated by detecting reverse transcriptase activity and by carrying out tests for detecting the p24 antigen (Elavia p24 polyclonal, SDP) over a period of one month.

5) Sequences:

The PCR products are visualized on agarose gels of from 1 to 1.4% concentration, depending on the size of the fragments, precipitated in 3M sodium acetate (1:10) and 3 volumes of absolute ethanol, incubated at -80°C for 30 minutes and then centrifuged at 13,000 rpm for 20 minutes. The pellet is dried and then taken up in 10 µl of distilled water (Sigma). Purification is carried out on a "Qiaquick Gel Extraction kit" (Qiagen) in accordance with the manufacturer's instructions; the



products are sequenced on an automated DNA sequencer (Applied Biosystems, Inc., Foster City, CA) using an Applied Biosystem Dye Terminator kit, as previously described (Loussert-Ajaka et al., 1995); the nucleotide
5 sequences are analysed on Sequence Navigator software (Applied Biosystems), and aligned using GeneWorks software (Intelligenetics Inc.).

6) Phylogenetic analyses:

The sequences were aligned using the CLUSTAL
10 software for multiple alignments and taking, as the reference matrix, the alignments of the compilation of HIV sequences possessed by the Laboratory of Biology and Theoretical Biophysics, Los Alamos, New Mexico, 87545 USA.

15 The phylogenetic analyses were performed using the PHYLIP software; the distances were firstly calculated using DNADIST, after which the phylogenetic analysis was carried out using NEIGHBOR JOINING or FITCH; finally, the trees were drawn using DRAWTREE
20 (Figures 9 to 19). The genetic distance percentages are also shown in Figure 20.

SEQBOOT was first of all used for the "bootstrapping" analyses, followed by DNADIST and NEIGHBOR JOINING or FITCH. Finally, the bootstrap
25 values were obtained using CONSENS.

II - Results of the investigation for detecting group O and variant HIV viruses:

174 samples, out of 3193 samples found to be positive in the screening, were regarded as being
30 group O or group M with abnormal serological reactivity or as being variants.

III - Detection of a non-group O and non-group M sample exhibiting abnormal serological reactivity

The 174 sera which were HIV-1-positive by WB
35 (Western blot), but reactive with a CO/OD ratio of < 5 in the competitive EIA, were tested by differential LIA dot blot on the V3 peptides from group M, group O and CPZGAB SIV:



- 7 do not react with any of the peptides represented (M, O or CPZGAB SIV). The absence of any cell collection does not allow any conclusion to be drawn.

5 - 82 give a reactivity with regard to at least one of the peptides corresponding to the V3 loop of O group strains. The frequency of the crossreactions is low and restricted to the epitopes which correspond to the consensus V3 regions (11%) and to the CPZGAB SIV V3
10 regions (43%).

- 84 sera do not react with the O group epitopes. Most of these samples were obtained from patients exhibiting an AIDS syndrome (75/84).

- one serum, which was taken from a Cameroonian
15 patient (NJ) reacts exclusively with the CPZGAB SIV peptide. This isolated reactivity with regard to a CPZGAB SIV antigen has never been described previously. Since lymphocytes had been collected from the patient, it was possible to continue with the virological
20 characterization of this strain, which was termed YBF30.

IV - Results of the serological and virological examinations performed on the first samples taken from this patient (May 1995) (serum No.: 95-6295):

25 1) Commercial ELISA tests (optical density/threshold value)

Criterion of positivity: $OD/CO > 1$

Génélatia = > 15

Wellcozyme $CO/OD = 1.55$

30 Abbott Plus = > 15

Behring Plus = 4.2

2) Western blot

New Lav 1 Pasteur WB:

160++, 120++, 68++, 55+, 41+, 40+/-, 34++,
35 24++, 18+

3) Innogenetics LIA dot blot

Negative for all the group O and group M bands
apart from CPZGAB SIV V3



4) Results of the investigative serological examinations carried out on peptides which are specific for the M and O groups

The technique developed by Professor Francis Barin of the Virology Laboratory of the Tours CHU was modified (Barin F. et al., 1996); use was made of synthesized transmembrane region peptides (BioMérieux) for developing a test for differentiating between the M and O groups. This technique is based on antibody-binding competition between the transmembrane gp41 peptides of the O and M groups, which are deposited on the solid phase, and gp41 transmembrane peptides either of the O group or of the M group at higher concentration in a hyperosmolar liquid reaction phase. The results are shown in Table I below, in which the CP well corresponds to the 100% inhibition control and the CSP well corresponds to the 0% inhibition control.

Table I

Results of the inter-group O-group M differentiations for the 6295 serum

	gp41 M	gp41 O	CP	CSP
6295	0.25	0.36	0.12	1.98

These results demonstrate that there is strong binding with regard to the peptides of the solid phase (CSP) and a marked inhibition due to the combined addition of the M and O peptides (CP), but no clear differentiation either by the M peptide or by the O peptide. This is, therefore, serological evidence that the infecting strain does not belong either to the M group or to the O group.

In view of an isolated reactivity in the InnoLia dot blot with regard to the CPZGAB SIV V3 antigens, on the same bases of competition between peptides, this serum was studied by bringing into competition the gp41 M, gp41 O and gp41 CPZGAB SIV peptides.

Use of the serum from the chimpanzee named 'Amandine' (donated by M. Peeters, who isolated the



CPZGAB SIV strain, AIDS 1992) initially enabled this technique to be validated. In Table II, the lowest values (OD) indicate the highest degree of binding to the antigens.

5

Table II

Results of the inter-group O-group M-CPZGAB SIV differentiations using the Amandine chimpanzee serum and the 6295 serum

	gp41 M	gp41 O	gp41 CPZGAB	CP	CSP
Amandine	0.8	1.4	0.3	0.5	1.9
6295	0.7	1.1	0.7	0.4	2.1

10

The reactivity of the "Amandine" serum confirms and validates the test according to the invention and shows that, while the serum of the patient reacts identically with regard to the M and CPZGAB SIV peptides, it does not exhibit a crossreaction with the O peptide.

15

These results demonstrate that the group M gp41 and CPZGAB SIV gp41 peptides exert a similar inhibition on the serum of the patient. The antigens of the infecting strain have therefore given rise to antibodies which recognize the group M and CPZGAB SIV gp41 peptides in a similar manner.

20

4) Results obtained from the lymphocyte isolation (sampling of May 1995)

A retrovirus was isolated, using standard techniques, from the lymphocytes which were sampled on 22 May 1995. Culture using the MT2 cell line shows that the YBF30 strain does not form any syncytia (NSI).

25

V - Results of the serological examinations carried out on the second blood sample (November 1995) (serum No.

30

95-3371)

1) Innogenetics LIA dot blot

Negative for all the bands, apart from CPZGAB SIV V3



2) Results of the investigative serological examinations carried out on the peptides specific for the M and O groups.

Table III shows the results of the inter-group
5 O-group M-CPZGAB SIV gp41 differentiations using the 3371 serum.

Table III

Results of the inter-group O-group M-CPZGAB SIV gp41
differentiations using the 3371 serum

	gp41 M	gp41 O	gp41 CPZGAB	CP	CSP
3371	1.31	1.7	0.89	0.54	2.02

10

These results confirm, on this new blood sample (taken from the same patient in the terminal stage of the disease), that the CPZGAB SIV gp41 peptide markedly inhibits the serum of the patient.

15 The antigens of the infecting strain have therefore induced antibodies which preferentially recognize the CPZGAB SIV gp41 peptide.

3) Results from the lymphocyte isolation (blood sampling of November 95 (95-3371-YBF31))

20 A retrovirus was isolated, using the standard techniques, from the lymphocytes which were sampled in November 1995 and termed YBF31; the sequence elements are identical to those of YBF30.

VI - Genomic amplification and sequences of YBF30

25 The DNA for all the PCR manipulations is extracted from the cells obtained at the end of a positive culture.

The PCRs carried out using the O group HIV-1 primers are negative in the different regions tested
30 (gag, pol, env). Similarly, those carried out using the primers which are specific for M group HIV-1 are also negative.

The amplification and hybridization conditions for the O group PCRs are those described in Loussert-Ajaka, 1995. The amplification and hybridization
35 conditions for the M group PCRs are those described by the authors cited below.



These M group primers are located in accordance with the HIV-1 HXB2 sequence as follows:

- in env gp120: ED3/ED12 (position 5956-5985; 7822-7792); ED5/ED14 (6556-6581; 7960-7931); ED5/ED12;
5 ED3/ED14; ES7/ES8 (7001-7020; 7667-7647) (Delwart et al. Science 1993; 262: 1257-1261).

- in env gp41: first PCR, ED3/M29, followed by a nested PCR, M28/M29 (7785-7808; 8099-8124); M28/M29 have the following sequences:

10 M28: CGGTTCTT(AG)GGAGCAGC(ACT)GGAAGCA,
M29: T(CT)T(ACGT)TCCCA(CT)T(AT)(CT)A(AGT)CCA(AGT)GTCAT;
SK68/SK69 (Ou et al. Science, 1988; 239: 295-297).

- in gag: Amplicor Roche Diagnostics systems;
15 nested gag primers (Loussert-Ajaka et al. Lancet 1995; 346: 912-913); SK38/SK39 (Ou et al., Science, 1988; 239: 295-297).

- in pol: A/NE1 (Boucher et al., Lancet, 1990; 336: 585-590); Pol3/Pol4 (Lauré et al., Lancet, 1988,
20 ii, 538-541).

Only the PCRs carried out using the H Pol primers (4235/4538) are positive, with this being followed by a nested PCR using the primers 4327/4481 (Fransen et al., Molecular and Cellular Probes 1994; 8:
25 317-322). This H Pol fragment, which is located in the integrase (260 bp), has been sequenced. Amplification using the HPOL primers is made possible due to the excess of virus. This is because the DNA which is used is extracted from cells at the end of a strongly
30 positive culture (reverse transcriptase > 100,000 cpm). It is not possible to amplify the DNA which is extracted from fresh cells without coculture because of the large number of mispairings between the HPOL primers (especially in the 3' region) and the sequence
35 of the YBF30 isolate. Conservation of this 3' end is very important for the extension activity of the Taq polymerase.

1 - Sequence of the pol gene: the use of very degenerate primers for amplifying, by RT-PCR, the RNA



extracted from the positive culture supernatant gave a positive amplification. These are primers which are common to all retroviruses (Donehower et al. J. Virol. Methods 1990; 28: 33-46), and are located in the reverse transcriptase region of the pol gene. Analysis of the fragment after sequencing made it possible to generate a specific primer, i.e. YRT2 (SEQ ID No.32), from the YBF30 isolate and to amplify the pol gene using the Hpol 4481 primer (Fransen et al., 1994, loc. cit.) as the antisense primer. The fragment was sequenced by synthesizing specific primers as required for each fragment generated (Figure 1).

2 - Sequence of the env gene: the second approach was to perform a long PCR (XL-PCR, Perkin Elmer), thereby amplifying all the virus (9000 bp) using primers situated in the LTR: LPBS 1 (SEQ ID No.22); LSiGi, followed by a 6000 bp nested PCR using YRT2 (SEQ ID No.32)/SK69, and to sequence all the envelope following the same procedure. The gp41 region was sequenced using a nested PCR and employing the primers SK68/LSiGi.

3 - Sequence of the gag gene: use of a nested PCR, achieved by means of a long PCR (LPBS 1/LSiGi), employing the primers Gag 5 and Gag 11i, and generating from this specific primers, as required, in order to walk along the viral genome.

VII - Results of the sequencings

The strain YBF30 was sequenced completely (see list of sequences). The YBF31 strain of November 1995 was sequenced in part, and the absence of significant variation confirms the validity of the YBF30 sequences.

VIII - Synthesizing peptides of the V3 loop region of the YBF30 strain.

Studying the sequences of the V3 loop region made it possible to synthesize the corresponding peptide and to compare the amino acids of this region of the YBF30 strain with those of other M subtypes and O strains.

The sequences of the peptides are:



- 15 -

YBF30: SEQ ID No.58
 CPZGAB SIV: CHRPGNNTRGEVQIGPGMTFYNIENVYGDTRSAYC
 (SEQ ID No.59)
 GROUP O: CIRPGNRTYRNLQIGPGMTFYNVEIATGDIRKAFC
 5 (ANT70) (SEQ ID No.60)
 GROUP M: CTRPNNNTRKSVRIGPGQAFYATGDIIGDIRQAHC
 (SS-TYPE A) (SEQ ID No.61)

The peptide was synthesized, starting with the two asparagines of the 5' region of the loop, and used in accordance with the same principle as previously described (see IV 4)), namely in competition in relation to the peptides of the M group, the O group and CPZGAB SIV. The results shown in Table IV confirm the original nature of this strain and the possible spread of these strains, since the serological results favour infection of the YBF30 type in Cameroon. Furthermore, a study of 200 selected HIV-1-positive sera from Cameroon provides evidence of a new case exhibiting a profile which is similar to that of YBF30.

Table IV
Study of the reactivity of 200 sera

Serum	Origin	V3A	V3cpz	V3YBF30	CP	CSP
953371	Cameroon	1.66	0.38	1.39	0.39	1.64
956295	Cameroon	1.72	0.37	1.16	0.51	1.73
967321	Cameroon	0.07	0.17	0.5	0.05	0.27
Amandine	GABSIV	1.74	0.14	1.48	0.19	1.74
NOA.*	ANTSIV	2.66	0.31	1.88	0.46	1.9

* serum from CPZ ANT SIV

The reactivity of the sera 953371 and 956295, corresponding to the patient from whom the YBF30 strain was isolated, with the CPZ SIV peptide, was confirmed in this new test. The lower reactivity with regard to its own V3 antigen is usual during the late stages of the disease. Nevertheless, this reactivity remains greater than that raised with regard to the M peptide. Another Cameroonian patient (serum 967321) exhibits the same profile of peptide reactivity.



References:

- 5 * Barin F. et al., *Aids Research and Human Retroviruses*, 1996, 12, 13, 1279-1289, *Diversity of Antibody Binding to V3 Peptides Representing Consensus Sequences of HIV Type 1 Genotypes A to E: An Approach for HIV Type 1 Serological Subtyping*.
- 10 * Charneau P., Borman AM., Quillent C., Guétard D., Chamaret S., Cohen J., Rémy G., Montagnier L., and F. Clavel, *Virology*, 1994, 205, 247-253, *Isolation and envelope sequence of a highly divergent HIV-1 isolate: definition of a new HIV-1 group*.
- 15 * Descamps D., Collin G., Lousert-Ajaka I., Saragosti S., Simon F. and F. Brun-Vezinet. *AIDS*, 1995, 9, 977-978, *HIV-1 group O sensitivity to antiretroviral drugs*.
- 20 * Huet, T., Cheynier R., Meyerhans A., Roelants G., and S. Wain-Hobson, *Nature*, 1990, 345, 356-359, *Genetic organization of a chimpanzee lentivirus related to HIV-1*.
- * Korber BTM., MacInnes K., Smith R. and G. Myers, *J. Virol.*, 1994, 68, 6730-6744, *Mutational trends in V3 loop protein sequences observed in different genetic lineages of HIV-1*.
- 25 * Lousert-Ajaka I., Ly TD., Chaix ML, Ingrand D., Saragosti S., Couroucé AM., Brun-Vezinet F. and F. Simon, *Lancet*, 1994, 343, 1393-1394, *HIV-1/HIV-2 seronegativity in HIV-1 subtype O infected patients*.
- 30 * Lousert-Ajaka I., Chaix ML., Korber B., Letourneur F., Gomas E., Allen E., Ly TD., Brun-Vezinet F., Simon F. and S. Saragosti, *J. Virol.*, 1995, 69, 5640-5649, *Variability of HIV type 1 group O strains isolated from Cameroonian patients living in FRANCE*.
- 35 * Murphy, E., B. Korber, Georges-Courbot, MC., You B., Pinter A., Cook D., Kienky MP., Georges A., Mathiot C., Barré-Sinoussi F., and M. Girard, *AIDS Res. Hum. Retroviruses*, 1993, 9, 997-1006,



Diversity of V3 region sequences of human immunodeficiency viruses type 1 from the Central African Republic.

- 5 * G. Myers, Aids Res. Hum. Retrovir., 1994, 10, 11, 1317-1324, Tenth Anniversary Perspectives on AIDS.
- * Nkengasong, J.N., Janssens W., Heyndrickx L., Fransen K., Ndumbe PM., Motte J., Leonaers A., Ngolle M., Ayuk J., Piot P., and G. Van der Groen, AIDS, 1994, 8, 1405-1412, Genotypic subtypes of HIV-1 in Cameroon.
- 10 * Sharp P.M. et al., AIDS, 1994, 8, suppl. 1, S27-S42, Origins and diversity of human immunodeficiency viruses.
- * Simon, F., T.D. Ly, A. Baillou-Beaufils, V. Schneider-Fauveau, J. de Saint-Martin, I. Loussert-Ajaka, M.L. Chaix, S. Saragosti, A.M. Couroucé, D. Ingrand, C. Janot, and F. Brun-Vezinet. AIDS, 1994, 8, 1628-1629. Sensitivity of screening kits for anti-HIV-1 subtype O antibodies.
- 20 * Zekeng, L., L. Gurtler, E. Afane Ze, A. Sam-Abbenyi, G. Mbouni, Essomba, E. Mpoudi-Ngolle, M. Monny-Lobbe, J.B. Tapko, and L. Kaptue, AIDS, 1994, 8, 1626-1628, Prevalence of HIV-1 subtype O infection in Cameroon: preliminary results.
- 25

 As is evident from the above, the invention is in no way limited to those of its embodiments which have just been described more explicitly; on the contrary, it encompasses all the variants which may come to the mind of the skilled person without departing from the context or scope of the present invention.

30



SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:

(A) NAME: INSTITUT NATIONAL DE LA SANTE ET DE LA
RECHERCHE MEDICALE - INSERM
(B) STREET: 101 rue DE TOLBIAC
(C) TOWN: PARIS
(E) COUNTRY: FRANCE
(F) ZIP CODE : 75654 CEDEX 13

(A) NAME: ASSISTANCE PUBLIQUE-HOPITAUX DE PARIS

(B) STREET: 3 avenue Victoria
(C) TOWN: PARIS
(E) COUNTRY: FRANCE
(F) ZIP CODE: 75100 RP

(A) NAME: INSTITUT PASTEUR
(B) STREET: 28 rue du Docteur Roux
(C) TOWN: PARIS
(E) COUNTRY: FRANCE
(F) ZIP CODE: 75724 Cédex 15

(A) NAME: MAUCLERE Philippe
(B) STREET: 2 rue Buhan
(C) TOWN: BORDEAUX
(E) COUNTRY: FRANCE
(F) ZIP CODE: 33000

(A) NAME: LOUSSET-AJAKA Ibtissam
(B) STREET: 26 avenue de la République
(C) TOWN: SARTROUVILLE
(E) COUNTRY: FRANCE
(F) ZIP CODE: 78500

(A) NAME: SIMON François
(B) STREET: 8 rue Germain Pilon
(C) TOWN: PARIS
(E) COUNTRY: FRANCE
(F) ZIP CODE: 75018

(A) NAME: SARAGOSTI Sentob
(B) STREET: 69 bis rue de Billancourt
(C) TOWN: BOULOGNE BILLANCOURT
(E) COUNTRY: FRANCE
(F) ZIP CODE: 92100

(A) NAME: BARRE-SINOUSSE Françoise
(B) STREET: 104 Le Capricorne, 50 rue d'Ereva
(C) TOWN: ISSY LES MOULINEAUX
(E) COUNTRY: FRANCE
(F) ZIP CODE: 92130

(ii) TITLE OF INVENTION: NON-M NON-O, HIV STRAINS,
FRAGMENTS AND USE.

(iii) NUMBER OF SEQUENCES: 61



(iv) COMPUTER READABLE FORM:

- (A) MEDIUM TYPE: Floppy disk
- (B) COMPUTER: IBM PC compatible
- (C) OPERATING SYSTEM : PC-DOS/MS-DOS
- (D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (OEB)

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGHT: 9183 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: simple
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: ADN (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

```

CTTCTCGCTT GTACTGGGTC TCTCTTGCTG GACCAGATTA GAGCCTGGGA GCTCTCTGGC   60
TAGCAGGGAA CCCACTGCTT AAGCCTCAAT AAAGCTTGCC TTGAGTGCTA AAGTGGTGTG   120
TGCCCATCCA TTCGGTAACT CTGGTACCTA GAGATCCCTC AGACCATCTA GACTGAGTGA   180
AAAATCTCTA GCAGTGGCGC CCGAACAGGG ACTTGAAAAC GAAAGTAGAA CCGGAGGCTG   240
AATCTCTCGA CGCAGGACTC GGCTCGTTGG TGCACACAGC GAGAGGCGAG GCGGCGGAAG   300
TGTGAGTACG CAATTTTGAC TGGCGGTGGC CAGAAAGTAG GAGAGAGGAT GGGTGCGAGA   360
GCGTCAGTGT TAACAGGGGG AAAATTAGAT CAATGGGAAT CAATTTATTT GAGACCAGGG   420
GGAAAGAAAA AATACAGAAT GAAACATTTA GTATGGGCAA GCAGGGAGCT GGAAAGATTG   480
GCTTGTAACC CAGGTCTCAT GGACACAGCG GACGGCTGTG CCAAGTTACT AAATCAATTA   540
GAACCAGCTC TCAAGACAGG GTCAGAAGAA CTGCGCTCTT TATATAACGC TCTAGCAGTT   600
CTTTATTGTG TCCATAGTAG GATACAGATA CACAACACAC AGGAAGCTTT GGACAAGATA   660
AAAGAGAAAC AGGAACAGCA CAAGCCCGAG CCAAAAAACC CAGAAGCAGG GGCAGCGGCA   720
GCAACTGATA GCAATATCAG TAGGAATTAT CCTCTAGTCC AGACTGCTCA AGGACAAATG   780
GTACATCAGC CGCTGACACC CAGAACCTTA AATGCTTGGG TGAAAGTGAT AGAGGAGAAG   840
GCCTTTAGTC CAGAAGTAAT ACCAATGTTT ATGGCCTTGT CAGAAGGGGC AACGCCCTCA   900
GATCTAAATA CTATGTTAAA TACAGTAGGG GGACATCAGG CAGCAATGCA GATGCTGAAG   960
GAAGTCATCA ATGAGGAAGC AGCAGACTGG GATAGGACAC ATCCAGTCCC TGTGGGACCA 1020
CTACCCCCAG GGCAACTGAG AGACCCTAGA GGAAGTGATA TAGCAGGAAC AACTAGCACC 1080

```



CTGGCAGAAC AGGTGGCTTG GATGACTGCT AATCCTCCTG TTCCAGTAGG AGATATTTAT 1140
 AGAAGATGGA TAGTCCTGGG GTTAAACAGA ATTGTGAGAA TGTATAGTCC TGTCAGCATT 1200
 CTAGAGATCA AACAAGGACC AAAAGAACCC TTCAGAGACT ATGTAGACAG GTTCTACAAA 1260
 ACTCTAAGAG CAGAGCAGGC AACACAGGAA GTAAAGAATT GGATGACAGA AACACTCTTA 1320
 GTACAAAATG CAAACCCAGA TTGTAAACAG CTCCTAAAAG CATTAGGGCC AGGAGCTACC 1380
 TTAGAAGAGA TGATGACGGC CTGCCAGGGA GTGGGGGGAC CAGCACATAA GGCAAGAGTG 1440
 CTAGCAGAGG CTATGTCACA GGTGCAGCAG CCAACAACCTA GTGTCTTTGC ACAAAGGGGA 1500
 AACTTTAAAG GCATAAGGAA ACCCATTAA TGTTCATT GTGGCAAAGA GGGCCATTTG 1560
 GCAAGAACT GTAAGGCCCC TAGAAGAGGA GGCTGTTGGA AGTGTGGGCA AGAAGGACAT 1620
 CAAATGAAAG ATTGTAAAAA TGAAGGAAGA CAGGCTAATT TTTTAGGGAA GAGCTGGTCT 1680
 CCCTTCAAAG GGAGACCAGG AAACCTCCCC CAGACAACAA CAAGGAAAGA GCCCACAGCC 1740
 CCGCCACTAG AGAGTTATGG GTTTCAGGAG GAGAAGAGCA CACAGGGGAA GGAGATGCAG 1800
 GAGAACCAGG AGAGGACAGA GAACTCTCTG TACCCACCTT TAACTTCCCT CAGATCACTC 1860
 TTTGGCAACG ACCCGTCATC ACAGTAAAAA TAGGGAAAGA AGTAAGAGAA GCTCTTTTAG 1920
 ATACAGGAGC TGATGATACA GTAATAGAAG AGCTACAATT AGAGGGAAAA TGGAAACCAA 1980
 AAATGATAGG AGGAATTGGA GGATTTATCA AAGTGAGACA ATATGATAAT ATAACAGTAG 2040
 ACATACAGGG AAGAAAAGCA GTTGGTACAG TATTAGTAGG ACCAACACCT GTTAATATTA 2100
 TAGGAAGAAA TCTTTTAACC CAGATTGGCT GTACTTTAAA TTTTCCAATA AGTCCTATTG 2160
 AAACCTGTACC AGTAAAATTA AAACCAGGAA TGGATGGCCC AAAGGTAAAA CAATGGCCTT 2220
 TGACAACAGA AAAAATAGAG GCATTAAGAG AAATTTGTAC AGAAATGGAA AAGGAAGGAA 2280
 AAATTTCTAG AATAGGGCCT GAGAATCCAT ATAACACTCC AATTTTGTCT ATAAAAAAGA 2340
 AAGATAGCAC TAAATGGAGA AAATTAGTAG ATTTACAGGA ATTAAATAAA AGGACCCAAG 2400
 ATTTTGGGA AGTGCAGCTA GGAATTCCAC ATCCAGCAGG ATTAAAGCAG AAAAAATCAG 2460
 TGACAGTTTT GGATGTAGGA GATGCTTATT TTTTCATGTCC CTTGGACAAA GATTTTAGAA 2520
 AGTATACAGC TTTTACCATA CCTAGTATAA ACAATGAGAC ACCTGGTATT AGATACCAGT 2580
 ATAATGTGCT GCCACAAGGC TGGAAAGGGT CACCAGCAAT TTTTCAGAGT ACAATGACAA 2640
 AAATTCTAGA ACCATTCAGA GAGAAACATC CAGAGATAAT CATTTACCAG TACATGGATG 2700
 AACCTCTATGT GGGATCTGAC TTAGAACTAG CACAACATAG AGAGGCAGTA GAAGACCTTA 2760



GAGATCATCT TTTGAAGTGG GGCTTTACGA CCCCTGACAA AAAACATCAG AAGGAACCCC 2820
 CGTTCCTCTG GATGGGATAT GAACTCCATC CAGACAAATG GACAGTCCAG CCAATAAAGT 2880
 TACCAGAAAA GGATGTATGG ACTGTCAATG ATATACAGAA ATTAGTAGGA AAGTTAAATT 2940
 GGGCAAGTCA GATCTATCCA GGAATCAGAG TAAAACAGCT CTGTAAATTA ATCAGAGGAA 3000
 CCAAAGCTTT GACAGAAGTA GTCAACTTTA CAGAAGAAGC AGAATTAGAA CTAGCAGAAA 3060
 ACAGGGAGAT ATTAAAAGAA CCCCTGCATG GAGTCTATTA TGACCCAGGA AAAGAATTAG 3120
 TAGCAGAAAT TCAAAAGCAA GGACAAGGTC AGTGGACATA TCAGATTTAT CAGGAGTTAC 3180
 ATAAAAATTT AAAAACAGGA AAGTATGCAA AAATGAGATC TGCCCATACT AATGATATAA 3240
 AACAGTTAGT TGAAGTGGTA AGGAAAGTGG CAACAGAAAG TATAGTAATT TGGGGAAAGA 3300
 CTCCTAAATT TAGATTACCA GTACAAAAGG AAGTGTGGGA GGCATGGTGG ACCGATCATT 3360
 GGCAAGCAAC TTGGATTCCT GAGTGGGAAT TTGTCAACAC TCCTCCCCTT GTAAAATTAT 3420
 GGTATCAGTT AGAAACAGAG CCAATCAGTG GGCAGAAAC TTTCTATGTA GATGGAGCAG 3480
 CTAATAGGGA AACAAAATTG GGAAAAGCAG GTTTTGTGAC AGATAGGGGA AGACAGAAAG 3540
 TGGTCTCTAT TGCAGACACC ACCAATCAAA AGGCTGAGTT ACAAGCTATC CTTATGGCCT 3600
 TACAAGAGTC AGGACGGGAT GTAAACATAG TCACTGACTC TCAGTATGCT ATGGGAATAA 3660
 TTCATTCACA GCCAGATAAA AGTGAATCAG AATTGGTGAG CCAAATAATA GAAGAGCTCA 3720
 TAAAAAAGGA AAGAGTTTAT CTCTCTTGGG TACCTGCACA TAAAGGTATT GGAGGAAATG 3780
 AGCAGGTAGA CAAATTAGTT AGCTCAGGAA TTAGAAAAAT ATTATTCCTA GATGGTATAG 3840
 AAAAAGCCCA AGAAGATCAT GACAGATATC ACAGCAATTG GAAAGCAATG GCCAGTGATT 3900
 TTAACCTACC CCCCATAGTG GCAAAAGAAA TAGTAGCCAG CTGTGACAAA TGCCAGCTAA 3960
 AAGGGGAAGC CATGCATGGA CAGGTCAATT GTAGTCCAGG AGTGTGGCAA TTAGATTGTA 4020
 CACACTTAGA GGGAAAAATC ATCCTTGTGG CGGTCCATGT GGCCAGTGGC TACTTAGAAG 4080
 CAGAAGTTAT TCCTGCAGAG ACAGGACAGG AAACAGCATA TTTTATTTTA AAGTTAGCTG 4140
 GAAGATGGCC AGTAAAAGTT ATACACACTG ATAATGGATC CAATTTCACT AGTGCCACTG 4200
 TAAAAGCAGC CTGTTGGTGG GCAAATATCA AACAGGAATT TGGGATACCC TACAATCCTC 4260
 AAAGTCAGGG AGCAGTAGAG TCCATGAATA AAGAATTAAA GAAAATTATA GGACAAATCA 4320
 GAGATCAAGC AGAACATCTA AAGACAGCAG TGCAAATGGC GGTTTTTCATT CACAATTTTA 4380



AAAGAAAGG GGGGATTGGG GGTACACTG CAGGGGAAAG AATAATAGAC ATAATAGCAA 4440
 CAGACATACA GACAACAAAT TTACAAACAC AAATTTTAAA AGTTCAAAAT TTTCGGGTTT 4500
 ATTACAGAGA CAGCAGAGAT CCCATTTGGA AAGGACCAGC CAAACTTCTG TGGAAAGGAG 4560
 AAGGGGCAGT GGTAATTCAA GATAACGGGG ATATAAAAGT AGTCCCACGT AGGAAAGCAA 4620
 AAATAATTAG GGATTATGGA AAACAGATGG CAGGTGATGG TTGTGTGGCA AGTGGACAGG 4680
 ATGAAAATCA GGAAATGGAA TAGCTTAGTA AAACATCATA TGTATGTGTC AAAAAAGGCA 4740
 AAAGGATGGT ATTATAGACA TCATTATGAA ACACATCACC CAAAAATAAG TTCAGAAGTA 4800
 CATATCCCAG TAGGTCAGGC AAGATTAGTG ACAGTCACCTT ATTGGGGGCT AACACAGGA 4860
 GAACAGTCTT GGCATCTAGG ACATGGAGTA TCCATAGAAT GGAGACTAAG AAAATACAAG 4920
 ACACAAGTTG ATCCTGAAAT GGCAGACAAG CTAATACATC TTCATTATTT TGATTGTTTT 4980
 ACAGCCTCTG CCATAAGGCA AGCGGTCTTA GGGAGACCAG TATTACCTAG GTGTGAATAT 5040
 CCAGCAGGGC ACAAACAGGT AGGCACCCTA CAATATCTAG CACTAACAGC CTGGGTGGGA 5100
 GCAAAGAAGA GAAAGCCACC CTTACCTAGT GTGACTAAGC TAACAGAAGA TAGATGGAAC 5160
 GAGCACCAGA AGATGCAGGG CCACAGAGGG AACCCTATAA TGAATGGGCA CTAGAATTAT 5220
 TAGAAGAATT AAAAAATGAA GCTGTGCGCC ATTTTCCAAG GATTGGCTA CATGGGTTAG 5280
 GACAACACAT CTATAACACA TATGGAGACA CCTGGGAGGG GGTAGAGGCA ATTATCAGGA 5340
 TACTACAACA ATTACTGTTT ATCCATTATA GGATTGGCTG CCAGCACAGC AGAATAGGGA 5400
 TCACTCCTCA AAGGAGAAGG AATGGAACCA GTAGATCCTA GATTAGAGCC CTGGAATCAT 5460
 CCAGGAAGCC AACCTAAAAC AGCTTGCAAT AATTGCTATT GTAAAAGATG TTGCTATCAC 5520
 TGCTTATATT GCTTCACAAA GAAAGGCTTA GGCATCTCAT ATGGCAGGAA GAAGCGGAGT 5580
 CAACGACGAA GAACTCCTCA GAGCAGTAAG AGTCATCAAG ATCTTATACC AGAGCAGTAA 5640
 GTAAAACCTG TATATATGCT GTCATTGGGA TTCATAGCGT TAGGAGCAGC AGTTAGCATA 5700
 GCAGTAATAG TCTGGGCATT ACTATATAGA GAATATAAGA AAATAAAATT GCAGGAAAAA 5760
 ATAAACACA TAAGACAGAG AATAAGAGAA AGAGAAGAAG ATAGTGGCAA TGAAAGTGAT 5820
 GGGGATGCAG AGTGGTTGGA TGGGGATGAA GAGTGGTTGG TTA CTCTTCT ATCTTCTAGT 5880
 AAGCTTGATC AAGGTAATTG GGTCTGAACA ACATTGGGTA ACAGTGACT ATGGGGTACC 5940
 AGTATGGAGA GAAGCAGAGA CAACTCTTTT CTGTGCTTCA GATGCTAAAG CCCATAGTAC 6000
 AGAGGCTCAC AACATCTGGG CCACACAAGC ATGTGTTCCCT ACTGATCCCA ATCCACAAGA 6060



AGTGCTATTA CCCAATGTAA CTGAAAAATT TAATATGTGG GAAAATAAAA TGGCAGACCA 6120
 AATGCAAGAG GATATTATCA GTCTGTGGGA ACAGAGCTTA AAGCCCTGTG TTAAATTAAC 6180
 CCCATTATGT GTAACATATGC TTTGTAACGA TAGCTATGGG GAGGAAAGGA ACAATACAAA 6240
 TATGACAACA AGAGAACCAG ACATAGGATA CAAACAAATG AAAAATTGCT CATTCAATGC 6300
 AACCCTGAG CTAACAGATA AAAAGAAGCA AGTTTACTCT CTGTTTTATG TAGAAGATGT 6360
 AGTACCAATC AATGCCTATA ATAAAACATA TAGGCTAATA AATTGTAATA CCACAGCTGT 6420
 GACACAAGCT TGTCTAAGA CTTCTTTGA GCCAATTCCA ATACATTACT GTGCACCACC 6480
 AGGCTTTGCC ATTATGAAAT GTAATGAAGG AAACCTTAGT GGAAATGGAA GCTGTACAAA 6540
 TGTGAGTACT GTACAATGCA CACATGGAAT AAAGCCAGTG ATATCCACTC AGTTAATCCT 6600
 AAATGGAAGC TTAAATACAG ATGGAATTGT TATTAGAAAT GATAGTCACA GTAATCTGTT 6660
 GGTGCAATGG AATGAGACAG TGCCAATAAA TTGTACAAGG CCAGGAAATA ATACAGGAGG 6720
 ACAGGTGCAG ATAGGACCTG CTATGACATT TTATAACATA GAAAAAATAG TAGGAGACAT 6780
 TAGACAAGCA TACTGTAATG TCTCTAAGA ACTATGGGAA CCAATGTGGA ATAGAACAAG 6840
 AGAGGAAATA AAGAAAATCC TGGGGAAAAA CAACATAACC TTCAGGGCTC GAGAGAGGAA 6900
 TGAAGGAGAC CTAGAAGTGA CACACTTAAT GTTCAATTGT AGAGGAGAGT TTTTCTATTG 6960
 TAACACTTCC AAATTATTTA ATGAGGAATT ACTTAACGAG ACAGGTGAGC CTATTACTCT 7020
 GCCTTGTTAGA ATAAGACAGA TTGTAAATTT GTGGACAAGG GTAGGAAAAG GAATTTATGC 7080
 ACCACCAATT CGGGGAGTTC TTAAGTGTAC CTCCAATATT ACTGGACTGG TTCTAGAATA 7140
 TAGTGGTGGG CCTGACACCA AGGAAACAAT AGTATATCCC TCAGGAGGAA ACATGGTTAA 7200
 TCTCTGGAGA CAAGAGTTGT ATAAGTACAA AGTAGTTAGC ATAGAACCCA TAGGAGTAGC 7260
 ACCAGGTAAA GCTAAAAGAC GCACAGTGAG TAGAGAAAAA AGAGCAGCCT TTGGACTAGG 7320
 TGCGCTGTTT CTTGGGTTTC TTGGAGCAGC AGGGAGCACT ATGGGCGCAG CGTCAATAAC 7380
 GCTGACGGTA CAGGCCCGGA CATTATTATC TGGGATAGTG CAACAGCAGA ATATTCTGTT 7440
 GAGAGCAATA GAGGCGCAAC AACATTTGTT GCAACTCTCA ATCTGGGGCA TTAAACAGCT 7500
 CCAGGCAAAA GTCCTTGCTA TAGAAAGATA CCTTAGGGAT CAGCAAATCC TAAGTCTATG 7560
 GGGCTGCTCA GGAAAAACAA TATGCTATAC CACTGTGCCT TGGAATGAGA CTTGGAGCAA 7620
 CAATACCTCT TATGATACAA TCTGGAATAA TTTAACCTGG CAACAATGGG ATGAGAAAGT 7680



AAGAACTAT TCAGGTGTCA TTTTGGACT TATAGAACAG GCACAAGAAC AACAGAACAC 7740
 AAATGAGAAA TCACTCTTGG AATTGGATCA ATGGGACAGT CTGTGGAGCT GGTTTGGTAT 7800
 TACAAAATGG CTGTGGTATA TAAAAATAGC TATAATGATA GTAGCAGGCA TTGTAGGCAT 7860
 AAGAATCATA AGTATAGTAA TAACTATAAT AGCAAGAGTT AGGCAGGGAT ATTCTCCCCT 7920
 TTCGTTGCAG ACCCTTATCC CAACAGCAAG GGGACCAGAC AGGCCAGAAG AAACAGAAGG 7980
 AGGCGTTGGA GAGCAAGACA GAGGCAGATC CGTGCGATTA GTGAGCGGAT TCTCAGCTCT 8040
 TGTCTGGGAG GACCTCCGGA ACCTGTTGAT CTTCTCTAC CACCGCTTGA CAGACTCACT 8100
 CTTGATACTG AGGAGGACTC TGGAACCTCT GGGACAGAGT CTCAGCAGGG GACTGCAACT 8160
 ACTGAATGAA CTCAGAACAC ACTTGTGGGG AATACTTGCA TATTGGGGAA AAGAGTTAAG 8220
 GGATAGTGCT ATCAGCTTGC TTAATACAAC AGCTATTGTA GTAGCAGAAG GAACAGATAG 8280
 GATTATAGAA TTAGCACAAA GAATAGGAAG GGGAAATATTA CACATACCTA GAAGAATCAG 8340
 ACAAGGCCTA GAAAGAGCAC TGATATAAGA TGGGAAAGAT TTGGTCAAAG AGCAGCCTAG 8400
 TAGGATGGCC AGAAATCAGA GAAAGAATGA GAAGACAAAC GCAAGAACCA GCAGTAGAGC 8460
 CAGCAGTAGG AGCAGGAGCA GCTTCTCAAG ATCTAGCTAA TCGAGGGGCC ATCACCATAA 8520
 GAAATACTAG AGACAATAAT GAAAGTATAG CTTGGCTAGA AGCACAAGAA GAAGAAGAGG 8580
 AAGTAGGCTT TCCAGTACGC CCTCAGGTAC CATTAAGGCC AATAACCTAT AAACAGGCTT 8640
 TTGATCTTTC CTTCTTTTAA AAAGATAAGG GGGGACTGGA AGGGCTAGTT TGGTCCAGAA 8700
 AAAGGCAAGA TATTCTAGAC CTCTGGATGT ATCACACACA AGGCATCCTC CCTGACTGGC 8760
 ATAACTACAC ACCAGGGCCA GGAATTAGAT ACCCCGTAAC CTTTGGATGG TGCTTCAAAC 8820
 TAGTACCATT GTCAGCTGAA GAAGTAGAAG AGGCTAATGA AGGAGACAAC AATGCCCTCT 8880
 TACACCCCAT ATGTCAACAT GGAGCAGATG ATGATCATAA AGAAGTGTTG GTGTGGCGAT 8940
 TTGACAGCTC CCTAGCAAGA AGACATGTAG CAAGAGAGCT GCATCCGGAG TTTTACAAGA 9000
 ACTGCTGACA AGGGACTTTA CTGCTGACAA GGGACTTTAT ACTTGGGGAC TTTCCGCCAG 9060
 GGACTTTCCA GGGAGGTGTG GTTGGGGGAG TGGCTTGCCC TCAGAGCTGC ATAAAAGCAG 9120
 CCGCTTCTCG CTTGTACTGG GTCTCTCTTG CTGGACCAGA TTAGAGTCTG GGAGCATATT 9180
 GGG 9183

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:



- (A) LENGHT: 813 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: simple
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: ADN (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

TTGGAAGGGC TAGTTTGGTC CAGAAAAAGG CAAGATATTC TAGACCTCTG GATGTATCAC	60
ACACAAGGCA TCCTCCCTGA CTGGCATAAC TACACACCAG GGCCAGGAAT TAGATACCCC	120
GTAACCTTTG GATGGTGCTT CAAACTAGTA CCATTGTCAG CTGAAGAAGT AGAAGAGGCT	180
AATGAAGGAG ACAACAATGC CCTCTTACAC CCCATATGTC AACATGGAGC AGATGATGAT	240
CATAAAGAAG TGTGGTGTG GCGATTGAC AGTCCCTAG CAAGAAGACA TGTAGCAAGA	300
GAGCTGCATC CGGAGTTTTA CAAGAACTGC TGACAAGGGA CTTTACTGCT GACAAGGGAC	360
TTTATACTTG GGGACTTTCC GCCAGGGACT TTCCAGGGAG GTGTGGTTGG GGGAGTGGCT	420
TGCCCTCAGA GCTGCATAAA AGCAGCCGCT TCTCGCTTGT ACTGGGTCTC TCTTGCTGGA	480
CTATACAGAT TAGAGCCTGG GAGCTCTCTG GCTAGCAGGG AACCCACTGC TTAAGCCTCA	540
ATAAATACAG CTTGCCTTGA GTGCTAAAGT GGTGTGTGCC CATCCATTCG GTAACCTCTGG	600
TACCTAGAGA ATCCCTCAGA CCATCTAGAC TGAGTGAAAA ATCTCTAGCA GTGGCGCCCCG	660
AACAGGGACT TAGTTGAAAA CGAAAGTAGA ACCGGAGGCT GAATCTCTCG ACGCAGGACT	720
CGGCTCGTTG GTGCACACAG CGAGAGGCGA GGCGGCGGAA GTGTGAGTAC GCAATTTTGA	780
CTGGCGGTGG CCAGAAAGTA GGAGAGAGGG AGG	813

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGHT: 1539 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: simple
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: ADN (genomic)

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..1536

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:



ATG GGT GCG AGA GCG TCA GTG TTA ACA GGG GGA AAA TTA GAT CAA TGG Met Gly Ala Arg Ala Ser Val Leu Thr Gly Gly Lys Leu Asp Gln Trp	48
1 5 10 15	
GAA TCA ATT TAT TTG AGA CCA GGG GGA AAG AAA AAA TAC AGA ATG AAA Glu Ser Ile Tyr Leu Arg Pro Gly Gly Lys Lys Lys Tyr Arg Met Lys	96
20 25 30	
CAT TTA GTA TGG GCA AGC AGG GAG CTG GAA AGA TTC GCT TGT AAC CCA His Leu Val Trp Ala Ser Arg Glu Leu Glu Arg Phe Ala Cys Asn Pro	144
35 40 45	
GGT CTC ATG GAC ACA GCG GAC GGC TGT GCC AAG TTA CTA AAT CAA TTA Gly Leu Met Asp Thr Ala Asp Gly Cys Ala Lys Leu Leu Asn Gln Leu	192
50 55 60	
GAA CCA GCT CTC AAG ACA GGG TCA GAA GAA CTG CGC TCT TTA TAT AAC Glu Pro Ala Leu Lys Thr Gly Ser Glu Glu Leu Arg Ser Leu Tyr Asn	240
65 70 75 80	
GCT CTA GCA GTT CTT TAT TGT GTC CAT AGT AGG ATA CAG ATA CAC AAC Ala Leu Ala Val Leu Tyr Cys Val His Ser Arg Ile Gln Ile His Asn	288
85 90 95	
ACA CAG GAA GCT TTG GAC AAG ATA AAA GAG AAA CAG GAA CAG CAC AAG Thr Gln Glu Ala Leu Asp Lys Ile Lys Glu Lys Gln Glu Gln His Lys	336
100 105 110	
CCC GAG CCA AAA AAC CCA GAA GCA GGG GCA GCG GCA GCA ACT GAT AGC Pro Glu Pro Lys Asn Pro Glu Ala Gly Ala Ala Ala Ala Thr Asp Ser	384
115 120 125	
AAT ATC AGT AGG AAT TAT CCT CTA GTC CAG ACT GCT CAA GGA CAA ATG Asn Ile Ser Arg Asn Tyr Pro Leu Val Gln Thr Ala Gln Gly Gln Met	432
130 135 140	
GTA CAT CAG CCG CTG ACA CCC AGA ACC TTA AAT GCT TGG GTG AAA GTG Val His Gln Pro Leu Thr Pro Arg Thr Leu Asn Ala Trp Val Lys Val	480
145 150 155 160	
ATA GAG GAG AAG GCC TTT AGT CCA GAA GTA ATA CCA ATG TTT ATG GCC Ile Glu Glu Lys Ala Phe Ser Pro Glu Val Ile Pro Met Phe Met Ala	528
165 170 175	
TTG TCA GAA GGG GCA ACG CCC TCA GAT CTA AAT ACT ATG TTA AAT ACA Leu Ser Glu Gly Ala Thr Pro Ser Asp Leu Asn Thr Met Leu Asn Thr	576
180 185 190	
GTA GGG GGA CAT CAG GCA GCA ATG CAG ATG CTG AAG GAA GTC ATC AAT Val Gly Gly His Gln Ala Ala Met Gln Met Leu Lys Glu Val Ile Asn	624
195 200 205	
GAG GAA GCA GCA GAC TGG GAT AGG ACA CAT CCA GTC CCT GTG GGA CCA Glu Glu Ala Ala Asp Trp Asp Arg Thr His Pro Val Pro Val Gly Pro	672
210 215 220	



CTA CCC CCA GGG CAA CTG AGA GAC CCT AGA GGA AGT GAT ATA GCA GGA Leu Pro Pro Gly Gln Leu Arg Asp Pro Arg Gly Ser Asp Ile Ala Gly 225 230 235 240	720
ACA ACT AGC ACC CTG GCA GAA CAG GTG GCT TGG ATG ACT GCT AAT CCT Thr Thr Ser Thr Leu Ala Glu Gln Val Ala Trp Met Thr Ala Asn Pro 245 250 255	768
CCT GTT CCA GTA GGA GAT ATT TAT AGA AGA TGG ATA GTC CTG GGG TTA Pro Val Pro Val Gly Asp Ile Tyr Arg Arg Trp Ile Val Leu Gly Leu 260 265 270	816
AAC AGA ATT GTG AGA ATG TAT AGT CCT GTC AGC ATT CTA GAG ATC AAA Asn Arg Ile Val Arg Met Tyr Ser Pro Val Ser Ile Leu Glu Ile Lys 275 280 285	864
CAA GGA CCA AAA GAA CCC TTC AGA GAC TAT GTA GAC AGG TTC TAC AAA Gln Gly Pro Lys Glu Pro Phe Arg Asp Tyr Val Asp Arg Phe Tyr Lys 290 295 300	912
ACT CTA AGA GCA GAG CAG GCA ACA CAG GAA GTA AAG AAT TGG ATG ACA Thr Leu Arg Ala Glu Gln Ala Thr Gln Glu Val Lys Asn Trp Met Thr 305 310 315 320	960
GAA ACA CTC TTA GTA CAA AAT GCA AAC CCA GAT TGT AAA CAG CTC CTA Glu Thr Leu Leu Val Gln Asn Ala Asn Pro Asp Cys Lys Gln Leu Leu 325 330 335	1008
AAA GCA TTA GGG CCA GGA GCT ACC TTA GAA GAG ATG ATG ACG GCC TGC Lys Ala Leu Gly Pro Gly Ala Thr Leu Glu Glu Met Met Thr Ala Cys 340 345 350	1056
CAG GGA GTG GGG GGA CCA GCA CAT AAG GCA AGA GTG CTA GCA GAG GCT Gln Gly Val Gly Gly Pro Ala His Lys Ala Arg Val Leu Ala Glu Ala 355 360 365	1104
ATG TCA CAG GTG CAG CAG CCA ACA ACT AGT GTC TTT GCA CAA AGG GGA Met Ser Gln Val Gln Gln Pro Thr Thr Ser Val Phe Ala Gln Arg Gly 370 375 380	1152
AAC TTT AAA GGC ATA AGG AAA CCC ATT AAA TGT TTC AAT TGT GGC AAA Asn Phe Lys Gly Ile Arg Lys Pro Ile Lys Cys Phe Asn Cys Gly Lys 385 390 395 400	1200
GAG GGC CAT TTG GCA AGA AAC TGT AAG GCC CCT AGA AGA GGA GGC TGT Glu Gly His Leu Ala Arg Asn Cys Lys Ala Pro Arg Arg Gly Gly Cys 405 410 415	1248
TGG AAG TGT GGG CAA GAA GGA CAT CAA ATG AAA GAT TGT AAA AAT GAA Trp Lys Cys Gly Gln Glu Gly His Gln Met Lys Asp Cys Lys Asn Glu 420 425 430	1296
GGA AGA CAG GCT AAT TTT TTA GGG AAG AGC TGG TCT CCC TTC AAA GGG Gly Arg Gln Ala Asn Phe Leu Gly Lys Ser Trp Ser Pro Phe Lys Gly 435 440 445 450	1344



435	440	445	
AGA CCA GGA AAC TTC CCC CAG ACA ACA ACA AGG AAA GAG CCC ACA GCC			1392
Arg Pro Gly Asn Phe Pro Gln Thr Thr Thr Arg Lys Glu Pro Thr Ala			
450	455	460	
CCG CCA CTA GAG AGT TAT GGG TTT CAG GAG GAG AAG AGC ACA CAG GGG			1440
Pro Pro Leu Glu Ser Tyr Gly Phe Gln Glu Glu Lys Ser Thr Gln Gly			
465	470	475	480
AAG GAG ATG CAG GAG AAC CAG GAG AGG ACA GAG AAC TCT CTG TAC CCA			1488
Lys Glu Met Gln Glu Asn Gln Glu Arg Thr Glu Asn Ser Leu Tyr Pro			
485	490	495	
CCT TTA ACT TCC CTC AGA TCA CTC TTT GGC AAC GAC CCG TCA TCA CAG			1536
Pro Leu Thr Ser Leu Arg Ser Leu Phe Gly Asn Asp Pro Ser Ser Gln			
500	505	510	
TAA			1539

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 512 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

Met Gly Ala Arg Ala Ser Val Leu Thr Gly Gly Lys Leu Asp Gln Trp			
1	5	10	15
Glu Ser Ile Tyr Leu Arg Pro Gly Gly Lys Lys Lys Tyr Arg Met Lys			
20	25	30	
His Leu Val Trp Ala Ser Arg Glu Leu Glu Arg Phe Ala Cys Asn Pro			
35	40	45	
Gly Leu Met Asp Thr Ala Asp Gly Cys Ala Lys Leu Leu Asn Gln Leu			
50	55	60	
Glu Pro Ala Leu Lys Thr Gly Ser Glu Glu Leu Arg Ser Leu Tyr Asn			
65	70	75	80
Ala Leu Ala Val Leu Tyr Cys Val His Ser Arg Ile Gln Ile His Asn			
85	90	95	
Thr Gln Glu Ala Leu Asp Lys Ile Lys Glu Lys Gln Glu Gln His Lys			
100	105	110	
Pro Glu Pro Lys Asn Pro Glu Ala Gly Ala Ala Ala Ala Thr Asp Ser			
115	120	125	



Asn Ile Ser Arg Asn Tyr Pro Leu Val Gln Thr Ala Gln Gly Gln Met		
130	135	140
Val His Gln Pro Leu Thr Pro Arg Thr Leu Asn Ala Trp Val Lys Val		
145	150	155 160
Ile Glu Glu Lys Ala Phe Ser Pro Glu Val Ile Pro Met Phe Met Ala		
	165	170 175
Leu Ser Glu Gly Ala Thr Pro Ser Asp Leu Asn Thr Met Leu Asn Thr		
	180	185 190
Val Gly Gly His Gln Ala Ala Met Gln Met Leu Lys Glu Val Ile Asn		
	195	200 205
Glu Glu Ala Ala Asp Trp Asp Arg Thr His Pro Val Pro Val Gly Pro		
	210	215 220
Leu Pro Pro Gly Gln Leu Arg Asp Pro Arg Gly Ser Asp Ile Ala Gly		
	225	230 235 240
Thr Thr Ser Thr Leu Ala Glu Gln Val Ala Trp Met Thr Ala Asn Pro		
	245	250 255
Pro Val Pro Val Gly Asp Ile Tyr Arg Arg Trp Ile Val Leu Gly Leu		
	260	265 270
Asn Arg Ile Val Arg Met Tyr Ser Pro Val Ser Ile Leu Glu Ile Lys		
	275	280 285
Gln Gly Pro Lys Glu Pro Phe Arg Asp Tyr Val Asp Arg Phe Tyr Lys		
	290	295 300
Thr Leu Arg Ala Glu Gln Ala Thr Gln Glu Val Lys Asn Trp Met Thr		
	305	310 315 320
Glu Thr Leu Leu Val Gln Asn Ala Asn Pro Asp Cys Lys Gln Leu Leu		
	325	330 335
Lys Ala Leu Gly Pro Gly Ala Thr Leu Glu Glu Met Met Thr Ala Cys		
	340	345 350
Gln Gly Val Gly Gly Pro Ala His Lys Ala Arg Val Leu Ala Glu Ala		
	355	360 365
Met Ser Gln Val Gln Gln Pro Thr Thr Ser Val Phe Ala Gln Arg Gly		
	370	375 380
Asn Phe Lys Gly Ile Arg Lys Pro Ile Lys Cys Phe Asn Cys Gly Lys		
	385	390 395 400
Glu Gly His Leu Ala Arg Asn Cys Lys Ala Pro Arg Arg Gly Gly Cys		
	405	410 415
Trp Lys Cys Gly Gln Glu Gly His Gln Met Lys Asp Cys Lys Asn Glu		



420	425	430
Gly Arg Gln Ala Asn Phe Leu Gly Lys Ser Trp Ser Pro Phe Lys Gly		
435	440	445
Arg Pro Gly Asn Phe Pro Gln Thr Thr Thr Arg Lys Glu Pro Thr Ala		
450	455	460
Pro Pro Leu Glu Ser Tyr Gly Phe Gln Glu Glu Lys Ser Thr Gln Gly		
465	470	475
		480
Lys Glu Met Gln Glu Asn Gln Glu Arg Thr Glu Asn Ser Leu Tyr Pro		
485	490	495
Pro Leu Thr Ser Leu Arg Ser Leu Phe Gly Asn Asp Pro Ser Ser Gln		
500	505	510

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3045 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: simple
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: ADN (genomic)

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..3042

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

TTT TTT AGG GAA GAG CTG GTC TCC CTT CAA AGG GAG ACC AGG AAA CTT	48
Phe Phe Arg Glu Glu Leu Val Ser Leu Gln Arg Glu Thr Arg Lys Leu	
515	520
525	
CCC CCA GAC AAC AAC AAG GAA AGA GCC CAC AGC CCC GCC ACT AGA GAG	96
Pro Pro Asp Asn Asn Lys Glu Arg Ala His Ser Pro Ala Thr Arg Glu	
530	535
540	
TTA TGG GTT TCA GGA GGA GAA GAG CAC ACA GGG GAA GGA GAT GCA GGA	144
Leu Trp Val Ser Gly Gly Glu Glu His Thr Gly Glu Gly Asp Ala Gly	
545	550
555	560
GAA CCA GGA GAG GAC AGA GAA CTC TCT GTA CCC ACC TTT AAC TTC CCT	192
Glu Pro Gly Glu Asp Arg Glu Leu Ser Val Pro Thr Phe Asn Phe Pro	
565	570
575	
CAG ATC ACT CTT TGG CAA CGA CCC GTC ATC ACA GTA AAA ATA GGG AAA	240
Gln Ile Thr Leu Trp Gln Arg Pro Val Ile Thr Val Lys Ile Gly Lys	
580	585
590	



GAA GTA AGA GAA GCT CTT TTA GAT ACA GGA GCT GAT GAT ACA GTA ATA Glu Val Arg Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val Ile 595 600 605	288
GAA GAG CTA CAA TTA GAG GGA AAA TGG AAA CCA AAA ATG ATA GGA GGA Glu Glu Leu Gln Leu Glu Gly Lys Trp Lys Pro Lys Met Ile Gly Gly 610 615 620	336
ATT GGA GGA TTT ATC AAA GTG AGA CAA TAT GAT AAT ATA ACA GTA GAC Ile Gly Gly Phe Ile Lys Val Arg Gln Tyr Asp Asn Ile Thr Val Asp 625 630 635 640	384
ATA CAG GGA AGA AAA GCA GTT GGT ACA GTA TTA GTA GGA CCA ACA CCT Ile Gln Gly Arg Lys Ala Val Gly Thr Val Leu Val Gly Pro Thr Pro 645 650 655	432
GTT AAT ATT ATA GGA AGA AAT CTT TTA ACC CAG ATT GGC TGT ACT TTA Val Asn Ile Ile Gly Arg Asn Leu Leu Thr Gln Ile Gly Cys Thr Leu 660 665 670	480
AAT TTT CCA ATA AGT CCT ATT GAA ACT GTA CCA GTA AAA TTA AAA CCA Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys Pro 675 680 685	528
GGA ATG GAT GGC CCA AAG GTA AAA CAA TGG CCT TTG ACA ACA GAA AAA Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Thr Glu Lys 690 695 700	576
ATA GAG GCA TTA AGA GAA ATT TGT ACA GAA ATG GAA AAG GAA GGA AAA Ile Glu Ala Leu Arg Glu Ile Cys Thr Glu Met Glu Lys Glu Gly Lys 705 710 715 720	624
ATT TCT AGA ATA GGG CCT GAG AAT CCA TAT AAC ACT CCA ATT TTT GCT Ile Ser Arg Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Ile Phe Ala 725 730 735	672
ATA AAA AAG AAA GAT AGC ACT AAA TGG AGA AAA TTA GTA GAT TTC AGG Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe Arg 740 745 750	720
GAA TTA AAT AAA AGG ACC CAA GAT TTT TGG GAA GTG CAG CTA GGA ATT Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly Ile 755 760 765	768
CCA CAT CCA GCA GGA TTA AAG CAG AAA AAA TCA GTG ACA GTT TTG GAT Pro His Pro Ala Gly Leu Lys Gln Lys Lys Ser Val Thr Val Leu Asp 770 775 780	816
GTA GGA GAT GCT TAT TTT TCA TGT CCC TTG GAC AAA GAT TTT AGA AAG Val Gly Asp Ala Tyr Phe Ser Cys Pro Leu Asp Lys Asp Phe Arg Lys 785 790 795 800	864
TAT ACA GCT TTT ACC ATA CCT AGT ATA AAC AAT GAG ACA CCT GGT ATT Tyr Thr Ala Phe Thr Ile Pro Ser Ile Asn Asn Glu Thr Pro Gly Ile	912



805	810	815	
AGA TAC CAG TAT AAT GTG CTG CCA CAA GGC TGG AAA GGG TCA CCA GCA Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro Ala 820 825 830			960
ATT TTT CAG AGT ACA ATG ACA AAA ATT CTA GAA CCA TTC AGA GAG AAA Ile Phe Gln Ser Thr Met Thr Lys Ile Leu Glu Pro Phe Arg Glu Lys 835 840 845			1008
CAT CCA GAG ATA ATC ATT TAC CAG TAC ATG GAT GAC CTC TAT GTG GGA His Pro Glu Ile Ile Ile Tyr Gln Tyr Met Asp Asp Leu Tyr Val Gly 850 855 860			1056
TCT GAC TTA GAA CTA GCA CAA CAT AGA GAG GCA GTA GAA GAC CTC AGA Ser Asp Leu Glu Leu Ala Gln His Arg Glu Ala Val Glu Asp Leu Arg 865 870 875 880			1104
GAT CAT CTT TTG AAG TGG GGC TTT ACG ACC CCT GAC AAA AAA CAT CAG Asp His Leu Leu Lys Trp Gly Phe Thr Thr Pro Asp Lys Lys His Gln 885 890 895			1152
AAG GAG CCC CCG TTC CTC TGG ATG GGA TAT GAA CTC CAT CCA GAC AAA Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp Lys 900 905 910			1200
TGG ACA GTC CAG CCA ATA AAG TTA CCA GAA AAG GAT GTA TGG ACT GTC Trp Thr Val Gln Pro Ile Lys Leu Pro Glu Lys Asp Val Trp Thr Val 915 920 925			1248
AAT GAT ATA CAG AAA TTA GTA GGA AAG TTA AAT TGG GCA AGT CAG ATC Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln Ile 930 935 940			1296
TAT CCA GGA ATC AGA GTA AAA CAG CTC TGT AAA TTA ATC AGA GGA GCC Tyr Pro Gly Ile Arg Val Lys Gln Leu Cys Lys Leu Ile Arg Gly Ala 945 950 955 960			1344
AGA GCT TTG ACA GAA GTA GTC AAC TTT ACA GAA GAA GCA GAA TTA GAA Arg Ala Leu Thr Glu Val Val Asn Phe Thr Glu Glu Ala Glu Leu Glu 965 970 975			1392
CTA GCA GAA AAC AGG GAG ATA TTA AAA GAA CCC CTG CAT GGA GTC TAT Leu Ala Glu Asn Arg Glu Ile Leu Lys Glu Pro Leu His Gly Val Tyr 980 985 990			1440
TAT GAC CCA GGA AAA GAA TTA GTA GCA GAA ATT CAA AAG CAA GGA CAA Tyr Asp Pro Gly Lys Glu Leu Val Ala Glu Ile Gln Lys Gln Gly Gln 995 1000 1005			1488
GGT CAG TGG ACA TAT CAG ATT TAT CAG GAG TTA CAT AAA AAT TTA AAA Gly Gln Trp Thr Tyr Gln Ile Tyr Gln Glu Leu His Lys Asn Leu Lys 1010 1015 1020			1536
AGA GGA AAG TAT GCA AAA ATG AGA TCT GCC CAT ACT AAT GAT ATA AAA			1584



Thr Gly Lys Tyr Ala Lys Met Arg Ser Ala His Thr Asn Asp Ile Lys	
1025	1030 1035 1040
CAG TTA GTT GAA GTG GTA AGG AAA GTG GCA ACA GAA AGT ATA GTA ATT	1632
Gln Leu Val Glu Val Val Arg Lys Val Ala Thr Glu Ser Ile Val Ile	
1045	1050 1055
TGG GGA AAG ACT CCT AAA TTT AGA TTA CCA GTA CAA AAG GAA GTG TGG	1680
Trp Gly Lys Thr Pro Lys Phe Arg Leu Pro Val Gln Lys Glu Val Trp	
1060	1065 1070
GAG GCA TGG TGG ACC GAT CAT TGG CAA GCA ACT TGG ATT CCT GAG TGG	1728
Glu Ala Trp Trp Thr Asp His Trp Gln Ala Thr Trp Ile Pro Glu Trp	
1075	1080 1085
GAA TTT GTC AAC ACT CCT CCC CTT GTA AAA TTA TGG TAT CAG TTA GAA	1776
Glu Phe Val Asn Thr Pro Pro Leu Val Lys Leu Trp Tyr Gln Leu Glu	
1090	1095 1100
ACA GAG CCA ATC AGT GGG GCA GAA ACT TTC TAT GTA GAT GGA GCA GCT	1824
Thr Glu Pro Ile Ser Gly Ala Glu Thr Phe Tyr Val Asp Gly Ala Ala	
1105	1110 1115 1120
AAT AGG GAA ACA AAA TTG GGA AAA GCA GGT TTT GTG ACA GAT AGG GGA	1872
Asn Arg Glu Thr Lys Leu Gly Lys Ala Gly Phe Val Thr Asp Arg Gly	
1125	1130 1135
AGA CAG AAA GTG GTC TCT ATT GCA GAC ACC ACC AAT CAA AAG GCT GAG	1920
Arg Gln Lys Val Val Ser Ile Ala Asp Thr Thr Asn Gln Lys Ala Glu	
1140	1145 1150
TTA CAA GCT ATC CTT ATG GCC TTA CAA GAG TCA GGA CGG GAT GTA AAC	1968
Leu Gln Ala Ile Leu Met Ala Leu Gln Glu Ser Gly Arg Asp Val Asn	
1155	1160 1165
ATA GTC ACT GAC TCT CAG TAT GCT ATG GGA ATA ATT CAT TCA CAG CCA	2016
Ile Val Thr Asp Ser Gln Tyr Ala Met Gly Ile Ile His Ser Gln Pro	
1170	1175 1180
GAT AAA AGT GAA TCA GAA TTG GTG AGC CAA ATA ATA GAA GAG CTC ATA	2064
Asp Lys Ser Glu Ser Glu Leu Val Ser Gln Ile Ile Glu Glu Leu Ile	
1185	1190 1195 1200
AAA AAG GAA AGA GTT TAT CTC TCT TGG GTA CCT GCA CAT AAA GGT ATT	2112
Lys Lys Glu Arg Val Tyr Leu Ser Trp Val Pro Ala His Lys Gly Ile	
1205	1210 1215
GGA GGA AAT GAG CAG GTA GAC AAA TTA GTT AGC TCA GGA ATT AGA AAA	2160
Gly Gly Asn Glu Gln Val Asp Lys Leu Val Ser Ser Gly Ile Arg Lys	
1220	1225 1230
ATA TTA TTC CTA GAT GGT ATA GAA AAA GCC CAA GAA GAT CAT GAC AGA	2208
Ile Leu Phe Leu Asp Gly Ile Glu Lys Ala Gln Glu Asp His Asp Arg	
1235	1240 1245



TAT CAC AGC AAT TGG AAA GCA ATG GCC AGT GAT TTT AAC TTA CCC CCC	2256
Tyr His Ser Asn Trp Lys Ala Met Ala Ser Asp Phe Asn Leu Pro Pro	
1250 1255 1260	
ATA GTG GCA AAA GAA ATA GTA GCC AGC TGT GAC AAA TGC CAG CTA AAA	2304
Ile Val Ala Lys Glu Ile Val Ala Ser Cys Asp Lys Cys Gln Leu Lys	
1265 1270 1275 1280	
GGG GAA GCC ATG CAT GGA CAG GTC AAT TGT AGT CCA GGA GTG TGG CAA	2352
Gly Glu Ala Met His Gly Gln Val Asn Cys Ser Pro Gly Val Trp Gln	
1285 1290 1295	
TTA GAT TGT ACA CAC TTA GAG GGA AAA ATC ATC CTT GTG GCG GTC CAT	2400
Leu Asp Cys Thr His Leu Glu Gly Lys Ile Ile Leu Val Ala Val His	
1300 1305 1310	
GTG GCC AGT GGC TAC TTA GAA GCA GAA GTT ATT CCT GCA GAG ACA GGA	2448
Val Ala Ser Gly Tyr Leu Glu Ala Glu Val Ile Pro Ala Glu Thr Gly	
1315 1320 1325	
CAG GAA ACA GCA TAT TTT ATT TTA AAG TTA GCT GGA AGA TGG CCA GTA	2496
Gln Glu Thr Ala Tyr Phe Ile Leu Lys Leu Ala Gly Arg Trp Pro Val	
1330 1335 1340	
AAA GTT ATA CAC ACT GAT AAT GGA TCC AAT TTC ACT AGT GCC ACT GTA	2544
Lys Val Ile His Thr Asp Asn Gly Ser Asn Phe Thr Ser Ala Thr Val	
1345 1350 1355 1360	
AAA GCA GCC TGT TGG TGG GCA AAT ATC AAA CAG GAA TTT GGG ATA CCC	2592
Lys Ala Ala Cys Trp Trp Ala Asn Ile Lys Gln Glu Phe Gly Ile Pro	
1365 1370 1375	
TAC AAT CCT CAA AGT CAG GGA GCA GTA GAG TCC ATG AAT AAA GAA TTA	2640
Tyr Asn Pro Gln Ser Gln Gly Ala Val Glu Ser Met Asn Lys Glu Leu	
1380 1385 1390	
AAG AAA ATT ATA GGA CAA ATC AGA GAT CAA GCA GAA CAT CTA AAG ACA	2688
Lys Lys Ile Ile Gly Gln Ile Arg Asp Gln Ala Glu His Leu Lys Thr	
1395 1400 1405	
GCA GTG CAA ATG GCG GTT TTC ATT CAC AAT TTT AAA AGA AAA GGG GGG	2736
Ala Val Gln Met Ala Val Phe Ile His Asn Phe Lys Arg Lys Gly Gly	
1410 1415 1420	
ATT GGG GGG TAC ACT GCA GGG GAA AGA ATA ATA GAC ATA ATA GCA ACA	2784
Ile Gly Gly Tyr Thr Ala Gly Glu Arg Ile Ile Asp Ile Ile Ala Thr	
1425 1430 1435 1440	
GAC ATA CAG ACA ACA AAT TTA CAA ACA CAA ATT TTA AAA GTT CAA AAT	2832
Asp Ile Gln Thr Thr Asn Leu Gln Thr Gln Ile Leu Lys Val Gln Asn	
1445 1450 1455	
TTT CGG GTT TAT TAC AGA GAC AGC AGA GAT CCC ATT TGG AAA GGA CCA	2880
Phe Arg Val Tyr Tyr Arg Asp Ser Arg Asp Pro Ile Trp Lys Gly Pro	
1460 1465 1470	



GCC AAA CTT CTG TGG AAA GGA GAA GGG GCA GTG GTA ATT CAA GAT AAC 2928
 Ala Lys Leu Leu Trp Lys Gly Glu Gly Ala Val Val Ile Gln Asp Asn
 1475 1480 1485

GGG GAT ATA AAA GTA GTC CCA CGT AGG AAA GCA AAA ATA ATT AGG GAT 2976
 Gly Asp Ile Lys Val Val Pro Arg Arg Lys Ala Lys Ile Ile Arg Asp
 1490 1495 1500

TAT GGA AAA CAG ATG GCA GGT GAT GGT TGT GTG GCA AGT GGA CAG GAT 3024
 Tyr Gly Lys Gln Met Ala Gly Asp Gly Cys Val Ala Ser Gly Gln Asp
 1505 1510 1515 1520

GAA AAT CAG GAA ATG GAA TAG 3045
 Glu Asn Gln Glu Met Glu
 1525

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1014 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

Phe Phe Arg Glu Glu Leu Val Ser Leu Gln Arg Glu Thr Arg Lys Leu
 1 5 10 15

Pro Pro Asp Asn Asn Lys Glu Arg Ala His Ser Pro Ala Thr Arg Glu
 20 25 30

Leu Trp Val Ser Gly Gly Glu Glu His Thr Gly Glu Gly Asp Ala Gly
 35 40 45

Glu Pro Gly Glu Asp Arg Glu Leu Ser Val Pro Thr Phe Asn Phe Pro
 50 55 60

Gln Ile Thr Leu Trp Gln Arg Pro Val Ile Thr Val Lys Ile Gly Lys
 65 70 75 80

Glu Val Arg Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val Ile
 85 90 95

Glu Glu Leu Gln Leu Glu Gly Lys Trp Lys Pro Lys Met Ile Gly Gly
 100 105 110

Ile Gly Gly Phe Ile Lys Val Arg Gln Tyr Asp Asn Ile Thr Val Asp
 115 120 125

Ile Gln Gly Arg Lys Ala Val Gly Thr Val Leu Val Gly Pro Thr Pro
 130 135 140



Val Asn Ile Ile Gly Arg Asn Leu Leu Thr Gln Ile Gly Cys Thr Leu
 145 150 155 160
 Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys Pro
 165 170 175
 Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Thr Glu Lys
 180 185 190
 Ile Glu Ala Leu Arg Glu Ile Cys Thr Glu Met Glu Lys Glu Gly Lys
 195 200 205
 Ile Ser Arg Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Ile Phe Ala
 210 215 220
 Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe Arg
 225 230 235 240
 Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly Ile
 245 250 255
 Pro His Pro Ala Gly Leu Lys Gln Lys Lys Ser Val Thr Val Leu Asp
 260 265 270
 Val Gly Asp Ala Tyr Phe Ser Cys Pro Leu Asp Lys Asp Phe Arg Lys
 275 280 285
 Tyr Thr Ala Phe Thr Ile Pro Ser Ile Asn Asn Glu Thr Pro Gly Ile
 290 295 300
 Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro Ala
 305 310 315 320
 Ile Phe Gln Ser Thr Met Thr Lys Ile Leu Glu Pro Phe Arg Glu Lys
 325 330 335
 His Pro Glu Ile Ile Ile Tyr Gln Tyr Met Asp Asp Leu Tyr Val Gly
 340 345 350
 Ser Asp Leu Glu Leu Ala Gln His Arg Glu Ala Val Glu Asp Leu Arg
 355 360 365
 Asp His Leu Leu Lys Trp Gly Phe Thr Thr Pro Asp Lys Lys His Gln
 370 375 380
 Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp Lys
 385 390 395 400
 Trp Thr Val Gln Pro Ile Lys Leu Pro Glu Lys Asp Val Trp Thr Val
 405 410 415
 Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln Ile
 420 425 430
 Tyr Pro Gly Ile Arg Val Lys Gln Leu Cys Lys Leu Ile Arg Gly Ala



435					440					445					
Arg	Ala	Leu	Thr	Glu	Val	Val	Asn	Phe	Thr	Glu	Glu	Ala	Glu	Leu	Glu
450					455					460					
Leu	Ala	Glu	Asn	Arg	Glu	Ile	Leu	Lys	Glu	Pro	Leu	His	Gly	Val	Tyr
465					470					475					480
Tyr	Asp	Pro	Gly	Lys	Glu	Leu	Val	Ala	Glu	Ile	Gln	Lys	Gln	Gly	Gln
485					490					495					
Gly	Gln	Trp	Thr	Tyr	Gln	Ile	Tyr	Gln	Glu	Leu	His	Lys	Asn	Leu	Lys
500					505					510					
Thr	Gly	Lys	Tyr	Ala	Lys	Met	Arg	Ser	Ala	His	Thr	Asn	Asp	Ile	Lys
515					520					525					
Gln	Leu	Val	Glu	Val	Val	Arg	Lys	Val	Ala	Thr	Glu	Ser	Ile	Val	Ile
530					535					540					
Trp	Gly	Lys	Thr	Pro	Lys	Phe	Arg	Leu	Pro	Val	Gln	Lys	Glu	Val	Trp
545					550					555					560
Glu	Ala	Trp	Trp	Thr	Asp	His	Trp	Gln	Ala	Thr	Trp	Ile	Pro	Glu	Trp
565					570					575					
Glu	Phe	Val	Asn	Thr	Pro	Pro	Leu	Val	Lys	Leu	Trp	Tyr	Gln	Leu	Glu
580					585					590					
Thr	Glu	Pro	Ile	Ser	Gly	Ala	Glu	Thr	Phe	Tyr	Val	Asp	Gly	Ala	Ala
595					600					605					
Asn	Arg	Glu	Thr	Lys	Leu	Gly	Lys	Ala	Gly	Phe	Val	Thr	Asp	Arg	Gly
610					615					620					
Arg	Gln	Lys	Val	Val	Ser	Ile	Ala	Asp	Thr	Thr	Asn	Gln	Lys	Ala	Glu
625					630					635					640
Leu	Gln	Ala	Ile	Leu	Met	Ala	Leu	Gln	Glu	Ser	Gly	Arg	Asp	Val	Asn
645					650					655					
Ile	Val	Thr	Asp	Ser	Gln	Tyr	Ala	Met	Gly	Ile	Ile	His	Ser	Gln	Pro
660					665					670					
Asp	Lys	Ser	Glu	Ser	Glu	Leu	Val	Ser	Gln	Ile	Ile	Glu	Glu	Leu	Ile
675					680					685					
Lys	Lys	Glu	Arg	Val	Tyr	Leu	Ser	Trp	Val	Pro	Ala	His	Lys	Gly	Ile
690					695					700					
Gly	Gly	Asn	Glu	Gln	Val	Asp	Lys	Leu	Val	Ser	Ser	Gly	Ile	Arg	Lys
705					710					715					720
Ile	Leu	Phe	Leu	Asp	Gly	Ile	Glu	Lys	Ala	Gln	Glu	Asp	His	Asp	Arg
725					730					735					



Tyr His Ser Asn Trp Lys Ala Met Ala Ser Asp Phe Asn Leu Pro Pro
 740 745 750
 Ile Val Ala Lys Glu Ile Val Ala Ser Cys Asp Lys Cys Gln Leu Lys
 755 760 765
 Gly Glu Ala Met His Gly Gln Val Asn Cys Ser Pro Gly Val Trp Gln
 770 775 780
 Leu Asp Cys Thr His Leu Glu Gly Lys Ile Ile Leu Val Ala Val His
 785 790 795 800
 Val Ala Ser Gly Tyr Leu Glu Ala Glu Val Ile Pro Ala Glu Thr Gly
 805 810 815
 Gln Glu Thr Ala Tyr Phe Ile Leu Lys Leu Ala Gly Arg Trp Pro Val
 820 825 830
 Lys Val Ile His Thr Asp Asn Gly Ser Asn Phe Thr Ser Ala Thr Val
 835 840 845
 Lys Ala Ala Cys Trp Trp Ala Asn Ile Lys Gln Glu Phe Gly Ile Pro
 850 855 860
 Tyr Asn Pro Gln Ser Gln Gly Ala Val Glu Ser Met Asn Lys Glu Leu
 865 870 875 880
 Lys Lys Ile Ile Gly Gln Ile Arg Asp Gln Ala Glu His Leu Lys Thr
 885 890 895
 Ala Val Gln Met Ala Val Phe Ile His Asn Phe Lys Arg Lys Gly Gly
 900 905 910
 Ile Gly Gly Tyr Thr Ala Gly Glu Arg Ile Ile Asp Ile Ile Ala Thr
 915 920 925
 Asp Ile Gln Thr Thr Asn Leu Gln Thr Gln Ile Leu Lys Val Gln Asn
 930 935 940
 Phe Arg Val Tyr Tyr Arg Asp Ser Arg Asp Pro Ile Trp Lys Gly Pro
 945 950 955 960
 Ala Lys Leu Leu Trp Lys Gly Glu Gly Ala Val Val Ile Gln Asp Asn
 965 970 975
 Gly Asp Ile Lys Val Val Pro Arg Arg Lys Ala Lys Ile Ile Arg Asp
 980 985 990
 Tyr Gly Lys Gln Met Ala Gly Asp Gly Cys Val Ala Ser Gly Gln Asp
 995 1000 1005

Glu Asn Gln Glu Met Glu
 1010



(A) LENGHT: 579 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: simple
(D) TOPOLOGY: linear

(ix) FEATURE:

(A) NAME/KEY: CDS
(B) LOCATION: 1..576

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

ATG GAA AAC AGA TGG CAG GTG ATG GTT GTG TGG CAA GTG GAC AGG ATG 48
Met Glu Asn Arg Trp Gln Val Met Val Val Trp Gln Val Asp Arg Met
1015 1020 1025 1030

AAA ATC AGG AAA TGG AAT AGC TTA GTA AAA CAT CAT ATG TAT GTG TCA 96
Lys Ile Arg Lys Trp Asn Ser Leu Val Lys His His Met Tyr Val Ser
1035 1040 1045

AAA AAG GCA AAA GGA TGG TAT TAT AGA CAT CAT TAT GAA ACA CAT CAC 144
Lys Lys Ala Lys Gly Trp Tyr Tyr Arg His His Tyr Glu Thr His His
1050 1055 1060

CCA AAA ATA AGT TCA GAA GTA CAT ATC CCA GTA GGT CAG GCA AGA TTA 192
Pro Lys Ile Ser Ser Glu Val His Ile Pro Val Gly Gln Ala Arg Leu
1065 1070 1075

GTG ACA GTC ACT TAT TGG GGG CTA ACA ACA GGA GAA CAG TCT TGG CAT 240
Val Thr Val Thr Tyr Trp Gly Leu Thr Thr Gly Glu Gln Ser Trp His
1080 1085 1090

CTA GGA CAT GGA GTA TCC ATA GAA TGG AGA CTA AGA AAA TAC AAG ACA 288
Leu Gly His Gly Val Ser Ile Glu Trp Arg Leu Arg Lys Tyr Lys Thr
1095 1100 1105 1110

CAA GTT GAT CCT GAA ATG GCA GAC AAG CTA ATA CAT CTT CAT TAT TTT 336
Gln Val Asp Pro Glu Met Ala Asp Lys Leu Ile His Leu His Tyr Phe
1115 1120 1125

GAT TGT TTT ACA GCC TCT GCC ATA AGG CAA GCG GTC TTA GGG AGA CCA 384
Asp Cys Phe Thr Ala Ser Ala Ile Arg Gln Ala Val Leu Gly Arg Pro
1130 1135 1140

GTA TTA CCT AGG TGT GAA TAT CCA GCA GGG CAC AAA CAG GTA GGC ACC 432
Val Leu Pro Arg Cys Glu Tyr Pro Ala Gly His Lys Gln Val Gly Thr
1145 1150 1155

CTA CAA TAT CTA GCA CTA ACA GCC TGG GTG GGA GCA AAG AAG AGA AAG 480
Leu Gln Tyr Leu Ala Leu Thr Ala Trp Val Gly Ala Lys Lys Arg Lys



1160	1165	1170	
CCA CCC TTA CCT AGT GTG ACT AAG CTA ACA GAA GAT AGA TGG AAC GAG			528
Pro Pro Leu Pro Ser Val Thr Lys Leu Thr Glu Asp Arg Trp Asn Glu			
1175	1180	1185	1190
CAC CAG AAG ATG CAG GGC CAC AGA GGG AAC CCT ATA ATG AAT GGG CAC			576
His Gln Lys Met Gln Gly His Arg Gly Asn Pro Ile Met Asn Gly His			
	1195	1200	1205
TAG			579

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 192 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

Met	Glu	Asn	Arg	Trp	Gln	Val	Met	Val	Val	Trp	Gln	Val	Asp	Arg	Met	
1				5					10					15		
Lys	Ile	Arg	Lys	Trp	Asn	Ser	Leu	Val	Lys	His	His	Met	Tyr	Val	Ser	
		20						25					30			
Lys	Lys	Ala	Lys	Gly	Trp	Tyr	Tyr	Arg	His	His	Tyr	Glu	Thr	His	His	
		35					40					45				
Pro	Lys	Ile	Ser	Ser	Glu	Val	His	Ile	Pro	Val	Gly	Gln	Ala	Arg	Leu	
		50					55					60				
Val	Thr	Val	Thr	Tyr	Trp	Gly	Leu	Thr	Thr	Gly	Glu	Gln	Ser	Trp	His	
	65				70					75					80	
Leu	Gly	His	Gly	Val	Ser	Ile	Glu	Trp	Arg	Leu	Arg	Lys	Tyr	Lys	Thr	
			85						90					95		
Gln	Val	Asp	Pro	Glu	Met	Ala	Asp	Lys	Leu	Ile	His	Leu	His	Tyr	Phe	
			100					105					110			
Asp	Cys	Phe	Thr	Ala	Ser	Ala	Ile	Arg	Gln	Ala	Val	Leu	Gly	Arg	Pro	
		115					120					125				
Val	Leu	Pro	Arg	Cys	Glu	Tyr	Pro	Ala	Gly	His	Lys	Gln	Val	Gly	Thr	
	130					135					140					
Leu	Gln	Tyr	Leu	Ala	Leu	Thr	Ala	Trp	Val	Gly	Ala	Lys	Lys	Arg	Lys	
145					150					155					160	
Pro	Pro	Leu	Pro	Ser	Val	Thr	Lys	Leu	Thr	Glu	Asp	Arg	Trp	Asn	Glu	
				165					170					175		



His Gln Lys Met Gln Gly His Arg Gly Asn Pro Ile Met Asn Gly His
 180 185 190

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 288 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: simple
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: ADN (genomic)

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..285

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

ATG GAA CGA GCA CCA GAA GAT GCA GGG CCA CAG AGG GAA CCC TAT AAT	48
Met Glu Arg Ala Pro Glu Asp Ala Gly Pro Gln Arg Glu Pro Tyr Asn	
195 200 205	
GAA TGG GCA CTA GAA TTA TTA GAA GAA TTA AAA AAT GAA GCT GTG CGC	96
Glu Trp Ala Leu Glu Leu Leu Glu Glu Leu Lys Asn Glu Ala Val Arg	
210 215 220	
CAT TTT CCA AGG ATT TGG CTA CAT GGG TTA GGA CAA CAC ATC TAT AAC	144
His Phe Pro Arg Ile Trp Leu His Gly Leu Gly Gln His Ile Tyr Asn	
225 230 235 240	
ACA TAT GGA GAC ACC TGG GAG GGG GTA GAG GCA ATT ATC AGG ATA CTA	192
Thr Tyr Gly Asp Thr Trp Glu Gly Val Glu Ala Ile Ile Arg Ile Leu	
245 250 255	
CAA CAA TTA CTG TTT ATC CAT TAT AGG ATT GGC TGC CAG CAC AGC AGA	240
Gln Gln Leu Leu Phe Ile His Tyr Arg Ile Gly Cys Gln His Ser Arg	
260 265 270	
ATA GGG ATC ACT CCT CAA AGG AGA AGG AAT GGA ACC AGT AGA TCC	285
Ile Gly Ile Thr Pro Gln Arg Arg Arg Asn Gly Thr Ser Arg Ser	
275 280 285	
TAG	288

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 95 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear



(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

```

Met Glu Arg Ala Pro Glu Asp Ala Gly Pro Gln Arg Glu Pro Tyr Asn
 1              5              10              15

Glu Trp Ala Leu Glu Leu Leu Glu Glu Leu Lys Asn Glu Ala Val Arg
      20              25              30

His Phe Pro Arg Ile Trp Leu His Gly Leu Gly Gln His Ile Tyr Asn
      35              40              45

Thr Tyr Gly Asp Thr Trp Glu Gly Val Glu Ala Ile Ile Arg Ile Leu
      50              55              60

Gln Gln Leu Leu Phe Ile His Tyr Arg Ile Gly Cys Gln His Ser Arg
      65              70              75              80

Ile Gly Ile Thr Pro Gln Arg Arg Arg Asn Gly Thr Ser Arg Ser
      85              90              95

```

(2) INFORMATION FOR SEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGHT: 252 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: simple
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: ADN (genomic)

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..249

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

```

ATG CTG TCA TTG GGA TTC ATA GCG TTA GGA GCA GCA GTT AGC ATA GCA      48
Met Leu Ser Leu Gly Phe Ile Ala Leu Gly Ala Ala Val Ser Ile Ala
      100              105              110

GTA ATA GTC TGG GCA TTA CTA TAT AGA GAA TAT AAG AAA ATA AAA TTG      96
Val Ile Val Trp Ala Leu Leu Tyr Arg Glu Tyr Lys Lys Ile Lys Leu
      115              120              125

CAG GAA AAA ATA AAA CAC ATA AGA CAG AGA ATA AGA GAA AGA GAA GAA      144
Gln Glu Lys Ile Lys His Ile Arg Gln Arg Ile Arg Glu Arg Glu Glu
      130              135              140

GAT AGT GGC AAT GAA AGT GAT GGG GAT GCA GAG TGG TTG GAT GGG GAT      192
Asp Ser Gly Asn Glu Ser Asp Gly Asp Ala Glu Trp Leu Asp Gly Asp
      145              150              155

```



GAA GAG TGG TTG GTT ACT CTT CTA TCT TCT AGT AAG CTT GAT CAA GGT 240
 Glu Glu Trp Leu Val Thr Leu Leu Ser Ser Ser Lys Leu Asp Gln Gly
 160 165 170 175

AAT TGG GTC TGA 252
 Asn Trp Val

(2) INFORMATION FOR SEQ ID NO: 12:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGHT: 83 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

Met Leu Ser Leu Gly Phe Ile Ala Leu Gly Ala Ala Val Ser Ile Ala
 1 5 10 15

Val Ile Val Trp Ala Leu Leu Tyr Arg Glu Tyr Lys Lys Ile Lys Leu
 20 25 30

Gln Glu Lys Ile Lys His Ile Arg Gln Arg Ile Arg Glu Arg Glu Glu
 35 40 45

Asp Ser Gly Asn Glu Ser Asp Gly Asp Ala Glu Trp Leu Asp Gly Asp
 50 55 60

Glu Glu Trp Leu Val Thr Leu Leu Ser Ser Ser Lys Leu Asp Gln Gly
 65 70 75 80

Asn Trp Val

(2) INFORMATION FOR SEQ ID NO: 13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGHT: 306 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: simple
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: ADN (genomic)

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..303

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

ATG GAA CCA GTA GAT CCT AGA TTA GAG CCC TGG AAT CAT CCA GGA AGC 48



Met	Glu	Pro	Val	Asp	Pro	Arg	Leu	Glu	Pro	Trp	Asn	His	Pro	Gly	Ser	
85						90					95					
CAA	CCT	AAA	ACA	GCT	TGC	AAT	AAT	TGC	TAT	TGT	AAA	AGA	TGT	TGC	TAT	96
Gln	Pro	Lys	Thr	Ala	Cys	Asn	Asn	Cys	Tyr	Cys	Lys	Arg	Cys	Cys	Tyr	
100					105					110					115	
CAC	TGC	TTA	TAT	TGC	TTC	ACA	AAG	AAA	GGC	TTA	GGC	ATC	TCA	TAT	GGC	144
His	Cys	Leu	Tyr	Cys	Phe	Thr	Lys	Lys	Gly	Leu	Gly	Ile	Ser	Tyr	Gly	
				120					125						130	
AGG	AAG	AAG	CGG	AGT	CAA	CGA	CGA	AGA	ACT	CCT	CAG	AGC	AGT	AAG	AGT	192
Arg	Lys	Lys	Arg	Ser	Gln	Arg	Arg	Arg	Thr	Pro	Gln	Ser	Ser	Lys	Ser	
			135					140					145			
CAT	CAA	GAT	CTT	ATA	CCA	GAG	CAG	CCC	TTA	TCC	CAA	CAG	CAA	GGG	GAC	240
His	Gln	Asp	Leu	Ile	Pro	Glu	Gln	Pro	Leu	Ser	Gln	Gln	Gln	Gly	Asp	
		150					155					160				
CAG	ACA	GGC	CAG	AAG	AAA	CAG	AAG	GAG	GCG	TTG	GAG	AGC	AAG	ACA	GAG	288
Gln	Thr	Gly	Gln	Lys	Lys	Gln	Lys	Glu	Ala	Leu	Glu	Ser	Lys	Thr	Glu	
	165					170					175					
GCA	GAT	CCG	TGC	GAT	TAG											306
Ala	Asp	Pro	Cys	Asp												
180																

(2) INFORMATION FOR SEQ ID NO: 14:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 101 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

Met	Glu	Pro	Val	Asp	Pro	Arg	Leu	Glu	Pro	Trp	Asn	His	Pro	Gly	Ser	
1				5					10					15		
Gln	Pro	Lys	Thr	Ala	Cys	Asn	Asn	Cys	Tyr	Cys	Lys	Arg	Cys	Cys	Tyr	
			20					25					30			
His	Cys	Leu	Tyr	Cys	Phe	Thr	Lys	Lys	Gly	Leu	Gly	Ile	Ser	Tyr	Gly	
		35					40					45				
Arg	Lys	Lys	Arg	Ser	Gln	Arg	Arg	Arg	Thr	Pro	Gln	Ser	Ser	Lys	Ser	
	50					55					60					
His	Gln	Asp	Leu	Ile	Pro	Glu	Gln	Pro	Leu	Ser	Gln	Gln	Gln	Gly	Asp	
65					70				75					80		
Gln	Thr	Gly	Gln	Lys	Lys	Gln	Lys	Glu	Ala	Leu	Glu	Ser	Lys	Thr	Glu	
			85					90						95		



Ala Asp Pro Cys Asp
100

(2) INFORMATION FOR SEQ ID NO: 15:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGHT: 369 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: simple
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: ADN (genomic)

- (ix) FEATURE:
 (A) NAME/KEY: CDS
 (B) LOCATION:1..366

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

ATG GCA GGA AGA AGC GGA GTC AAC GAC GAA GAA CTC CTC AGA GCA GTA	48
Met Ala Gly Arg Ser Gly Val Asn Asp Glu Glu Leu Leu Arg Ala Val	
105 110 115	
AGA GTC ATC AAG ATC TTA TAC CAG AGC AGT TAT CCC AAC AGC AAG GGG	96
Arg Val Ile Lys Ile Leu Tyr Gln Ser Ser Tyr Pro Asn Ser Lys Gly	
120 125 130	
ACC AGA CAG GCC AGA AGA AAC AGA AGG AGG CGT TGG AGA GCA AGA CAG	144
Thr Arg Gln Ala Arg Arg Asn Arg Arg Arg Arg Trp Arg Ala Arg Gln	
135 140 145	
AGG CAG ATC CGT GCG ATT AGT GAG CGG ATT CTC AGC TCT TGT CTG GGA	192
Arg Gln Ile Arg Ala Ile Ser Glu Arg Ile Leu Ser Ser Cys Leu Gly	
150 155 160 165	
GGA CCT CCG GAA CCT GTT GAT CTT CCT CTA CCA CCG CTT GAC AGA CTC	240
Gly Pro Pro Glu Pro Val Asp Leu Pro Leu Pro Pro Leu Asp Arg Leu	
170 175 180	
ACT CTT GAT ACT GAG GAG GAC TCT GGA ACT CCT GGG ACA GAG TCT CAG	288
Thr Leu Asp Thr Glu Glu Asp Ser Gly Thr Pro Gly Thr Glu Ser Gln	
185 190 195	
CAG GGG ACT GCA ACT ACT GAA TGA ACT CAG AAC ACA CTT GTG GGG AAT	336
Gln Gly Thr Ala Thr Thr Glu * Thr Gln Asn Thr Leu Val Gly Asn	
200 205 210	
ACT TGC ATA TTG GGG AAA AGA GTT AAG GGA TAG	369
Thr Cys Ile Leu Gly Lys Arg Val Lys Gly	
215 220	

(2) INFORMATION FOR SEQ ID NO: 16:



(i) SEQUENCE CHARACTERISTICS:

- (A) LENGHT: 122 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

```

Met Ala Gly Arg Ser Gly Val Asn Asp Glu Glu Leu Leu Arg Ala Val
 1             5             10             15
Arg Val Ile Lys Ile Leu Tyr Gln Ser Ser Tyr Pro Asn Ser Lys Gly
          20             25             30
Thr Arg Gln Ala Arg Arg Asn Arg Arg Arg Arg Trp Arg Ala Arg Gln
          35             40             45
Arg Gln Ile Arg Ala Ile Ser Glu Arg Ile Leu Ser Ser Cys Leu Gly
          50             55             60
Gly Pro Pro Glu Pro Val Asp Leu Pro Leu Pro Pro Leu Asp Arg Leu
65             70             75             80
Thr Leu Asp Thr Glu Glu Asp Ser Gly Thr Pro Gly Thr Glu Ser Gln
          85             90             95
Gln Gly Thr Ala Thr Thr Glu * Thr Gln Asn Thr Leu Val Gly Asn
          100             105             110
Thr Cys Ile Leu Gly Lys Arg Val Lys Gly
          115             120

```

(2) INFORMATION FOR SEQ ID NO: 17:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGHT: 2559 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: simple
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: ADN (genomic)

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION:1..2556

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

```

ATG AAA GTG ATG GGG ATG CAG AGT GGT TGG ATG GGG ATG AAG AGT GGT
Met Lys Val Met Gly Met Gln Ser Gly Trp Met Gly Met Lys Ser Gly
          125             130             135

```

48



TGG TTA CTC TTC TAT CTT CTA GTA AGC TTG ATC AAG GTA ATT GGG TCT	96
Trp Leu Leu Phe Tyr Leu Leu Val Ser Leu Ile Lys Val Ile Gly Ser	
140 145 150	
GAA CAA CAT TGG GTA ACA GTG TAC TAT GGG GTA CCA GTA TGG AGA GAA	144
Glu Gln His Trp Val Thr Val Tyr Tyr Gly Val Pro Val Trp Arg Glu	
155 160 165 170	
GCA GAG ACA ACT CTT TTC TGT GCT TCA GAT GCT AAA GCC CAT AGT ACA	192
Ala Glu Thr Thr Leu Phe Cys Ala Ser Asp Ala Lys Ala His Ser Thr	
175 180 185	
GAG GCT CAC AAC ATC TGG GCC ACA CAA GCA TGT GTT CCT ACT GAT CCC	240
Glu Ala His Asn Ile Trp Ala Thr Gln Ala Cys Val Pro Thr Asp Pro	
190 195 200	
AAT CCA CAA GAA GTG CTA TTA CCC AAT GTA ACT GAA AAA TTT AAT ATG	288
Asn Pro Gln Glu Val Leu Leu Pro Asn Val Thr Glu Lys Phe Asn Met	
205 210 215	
TGG GAA AAT AAA ATG GCA GAC CAA ATG CAA GAG GAT ATT ATC AGT CTG	336
Trp Glu Asn Lys Met Ala Asp Gln Met Gln Glu Asp Ile Ile Ser Leu	
220 225 230	
TGG GAA CAG AGC TTA AAG CCC TGT GTT AAA TTA ACC CCA TTA TGT GTA	384
Trp Glu Gln Ser Leu Lys Pro Cys Val Lys Leu Thr Pro Leu Cys Val	
235 240 245 250	
ACT ATG CTT TGT AAC GAT AGC TAT GGG GAG GAA AGG AAC AAT ACA AAT	432
Thr Met Leu Cys Asn Asp Ser Tyr Gly Glu Glu Arg Asn Asn Thr Asn	
255 260 265	
ATG ACA ACA AGA GAA CCA GAC ATA GGA TAC AAA CAA ATG AAA AAT TGC	480
Met Thr Thr Arg Glu Pro Asp Ile Gly Tyr Lys Gln Met Lys Asn Cys	
270 275 280	
TCA TTC AAT GCA ACC ACT GAG CTA ACA GAT AAA AAG AAG CAA GTT TAC	528
Ser Phe Asn Ala Thr Thr Glu Leu Thr Asp Lys Lys Lys Gln Val Tyr	
285 290 295	
TCT CTG TTT TAT GTA GAA GAT GTA GTA CCA ATC AAT GCC TAT AAT AAA	576
Ser Leu Phe Tyr Val Glu Asp Val Val Pro Ile Asn Ala Tyr Asn Lys	
300 305 310	
ACA TAT AGG CTA ATA AAT TGT AAT ACC ACA GCT GTG ACA CAA GCT TGT	624
Thr Tyr Arg Leu Ile Asn Cys Asn Thr Thr Ala Val Thr Gln Ala Cys	
315 320 325 330	
CCT AAG ACT TCC TTT GAG CCA ATT CCA ATA CAT TAC TGT GCA CCA CCA	672
Pro Lys Thr Ser Phe Glu Pro Ile Pro Ile His Tyr Cys Ala Pro Pro	
335 340 345	
GGC TTT GCC ATT ATG AAA TGT AAT GAA GGA AAC TTT AGT GGA AAT GGA	720
Gly Phe Ala Ile Met Lys Cys Asn Glu Gly Asn Phe Ser Gly Asn Gly	
350 355 360	



AGC TGT ACA AAT GTG AGT ACT GTA CAA TGC ACA CAT GGA ATA AAG CCA Ser Cys Thr Asn Val Ser Thr Val Gln Cys Thr His Gly Ile Lys Pro 365 370 375	768
GTG ATA TCC ACT CAG TTA ATC CTA AAT GGA AGC TTA AAT ACA GAT GGA Val Ile Ser Thr Gln Leu Ile Leu Asn Gly Ser Leu Asn Thr Asp Gly 380 385 390	816
ATT GTT ATT AGA AAT GAT AGT CAC AGT AAT CTG TTG GTG CAA TGG AAT Ile Val Ile Arg Asn Asp Ser His Ser Asn Leu Leu Val Gln Trp Asn 395 400 405 410	864
GAG ACA GTG CCA ATA AAT TGT ACA AGG CCA GGA AAT AAT ACA GGA GGA Glu Thr Val Pro Ile Asn Cys Thr Arg Pro Gly Asn Asn Thr Gly Gly 415 420 425	912
CAG GTG CAG ATA GGA CCT GCT ATG ACA TTT TAT AAC ATA GAA AAA ATA Gln Val Gln Ile Gly Pro Ala Met Thr Phe Tyr Asn Ile Glu Lys Ile 430 435 440	960
GTA GGA GAC ATT AGA CAA GCA TAC TGT AAT GTC TCT AAA GAA CTA TGG Val Gly Asp Ile Arg Gln Ala Tyr Cys Asn Val Ser Lys Glu Leu Trp 445 450 455	1008
GAA CCA ATG TGG AAT AGA ACA AGA GAG GAA ATA AAG AAA ATC CTG GGG Glu Pro Met Trp Asn Arg Thr Arg Glu Glu Ile Lys Lys Ile Leu Gly 460 465 470	1056
AAA AAC AAC ATA ACC TTC AGG GCT CGA GAG AGG AAT GAA GGA GAC CTA Lys Asn Asn Ile Thr Phe Arg Ala Arg Glu Arg Asn Glu Gly Asp Leu 475 480 485 490	1104
GAA GTG ACA CAC TTA ATG TTC AAT TGT AGA GGA GAG TTT TTC TAT TGT Glu Val Thr His Leu Met Phe Asn Cys Arg Gly Glu Phe Phe Tyr Cys 495 500 505	1152
AAC ACT TCC AAA TTA TTT AAT GAG GAA TTA CTT AAC GAG ACA GGT GAG Asn Thr Ser Lys Leu Phe Asn Glu Glu Leu Leu Asn Glu Thr Gly Glu 510 515 520	1200
CCT ATT ACT CTG CCT TGT AGA ATA AGA CAG ATT GTA AAT TTG TGG ACA Pro Ile Thr Leu Pro Cys Arg Ile Arg Gln Ile Val Asn Leu Trp Thr 525 530 535	1248
AGG GTA GGA AAA GGA ATT TAT GCA CCA CCA ATT CGG GGA GTT CTT AAC Arg Val Gly Lys Gly Ile Tyr Ala Pro Pro Ile Arg Gly Val Leu Asn 540 545 550	1296
TGT ACC TCC AAT ATT ACT GGA CTG GTT CTA GAA TAT AGT GGT GGG CCT Cys Thr Ser Asn Ile Thr Gly Leu Val Leu Glu Tyr Ser Gly Gly Pro 555 560 565 570	1344
GAC ACC AAG GAA ACA ATA GTA TAT CCC TCA GGA GGA AAC ATG GTT AAT Asp Thr Lys Glu Thr Ile Val Tyr Pro Ser Gly Gly Asn Met Val Asn	1392



575										580										585									
CTC	TGG	AGA	CAA	GAG	TTG	TAT	AAG	TAC	AAA	GTA	GTT	AGC	ATA	GAA	CCC														
Leu	Trp	Arg	Gln	Glu	Leu	Tyr	Lys	Tyr	Lys	Val	Val	Ser	Ile	Glu	Pro														1440
			590						595					600															
ATA	GGA	GTA	GCA	CCA	GGT	AAA	GCT	AAA	AGA	CGC	ACA	GTG	AGT	AGA	GAA														1488
Ile	Gly	Val	Ala	Pro	Gly	Lys	Ala	Lys	Arg	Arg	Thr	Val	Ser	Arg	Glu														
			605					610						615															
AAA	AGA	GCA	GCC	TTT	GGA	CTA	GGT	GCG	CTG	TTT	CTT	GGG	TTT	CTT	GGA														1536
Lys	Arg	Ala	Ala	Phe	Gly	Leu	Gly	Ala	Leu	Phe	Leu	Gly	Phe	Leu	Gly														
			620					625						630															
GCA	GCA	GGG	AGC	ACT	ATG	GGC	GCA	GCG	TCA	ATA	ACG	CTG	ACG	GTA	CAG														1584
Ala	Ala	Gly	Ser	Thr	Met	Gly	Ala	Ala	Ser	Ile	Thr	Leu	Thr	Val	Gln														
			635					640						645															
GCC	CGG	ACA	TTA	TTA	TCT	GGG	ATA	GTG	CAA	CAG	CAG	AAT	ATT	CTG	TTG														1632
Ala	Arg	Thr	Leu	Leu	Ser	Gly	Ile	Val	Gln	Gln	Gln	Asn	Ile	Leu	Leu														
						655								660															
AGA	GCA	ATA	GAG	GCG	CAA	CAA	CAT	TTG	TTG	CAA	CTC	TCA	ATC	TGG	GGC														1680
Arg	Ala	Ile	Glu	Ala	Gln	Gln	His	Leu	Leu	Gln	Leu	Ser	Ile	Trp	Gly														
						670								680															
ATT	AAA	CAG	CTC	CAG	GCA	AAA	GTC	CTT	GCT	ATA	GAA	AGA	TAC	CTT	AGG														1728
Ile	Lys	Gln	Leu	Gln	Ala	Lys	Val	Leu	Ala	Ile	Glu	Arg	Tyr	Leu	Arg														
						685								695															
GAT	CAG	CAA	ATC	CTA	AGT	CTA	TGG	GGC	TGC	TCA	GGA	AAA	ACA	ATA	TGC														1776
Asp	Gln	Gln	Ile	Leu	Ser	Leu	Trp	Gly	Cys	Ser	Gly	Lys	Thr	Ile	Cys														
						700								710															
TAT	ACC	ACT	GTG	CCT	TGG	AAT	GAG	ACT	TGG	AGC	AAC	AAT	ACC	TCT	TAT														1824
Tyr	Thr	Thr	Val	Pro	Trp	Asn	Glu	Thr	Trp	Ser	Asn	Asn	Thr	Ser	Tyr														
						715								720															
GAT	ACA	ATC	TGG	AAT	AAT	TTA	ACC	TGG	CAA	CAA	TGG	GAT	GAG	AAA	GTA														1872
Asp	Thr	Ile	Trp	Asn	Asn	Leu	Thr	Trp	Gln	Gln	Trp	Asp	Glu	Lys	Val														
						735								740															
AGA	AAC	TAT	TCA	GGT	GTC	ATT	TTT	GGA	CTT	ATA	GAA	CAG	GCA	CAA	GAA														1920
Arg	Asn	Tyr	Ser	Gly	Val	Ile	Phe	Gly	Leu	Ile	Glu	Gln	Ala	Gln	Glu														
						750								755															
CAA	CAG	AAC	ACA	AAT	GAG	AAA	TCA	CTC	TTG	GAA	TTG	GAT	CAA	TGG	GAC														1968
Gln	Gln	Asn	Thr	Asn	Glu	Lys	Ser	Leu	Leu	Glu	Leu	Asp	Gln	Trp	Asp														
						765								770															
AGT	CTG	TGG	AGC	TGG	TTT	GGT	ATT	ACA	AAA	TGG	CTG	TGG	TAT	ATA	AAA														2016
Ser	Leu	Trp	Ser	Trp	Phe	Gly	Ile	Thr	Lys	Trp	Leu	Trp	Tyr	Ile	Lys														
						780								790															
ATA	GCT	ATA	ATG	ATA	GTA	GCA	GGC	ATT	GTA	GGC	ATA	AGA	ATC	ATA	AGT														2064



Ile	Ala	Ile	Met	Ile	Val	Ala	Gly	Ile	Val	Gly	Ile	Arg	Ile	Ile	Ser	
795					800					805					810	
ATA	GTA	ATA	ACT	ATA	ATA	GCA	AGA	GTT	AGG	CAG	GGA	TAT	TCT	CCC	CTT	2112
Ile	Val	Ile	Thr	Ile	Ile	Ala	Arg	Val	Arg	Gln	Gly	Tyr	Ser	Pro	Leu	
			815					820						825		
TCG	TTG	CAG	ACC	CTT	ATC	CCA	ACA	GCA	AGG	GGA	CCA	GAC	AGG	CCA	GAA	2160
Ser	Leu	Gln	Thr	Leu	Ile	Pro	Thr	Ala	Arg	Gly	Pro	Asp	Arg	Pro	Glu	
			830					835					840			
GAA	ACA	GAA	GGA	GGC	GTT	GGA	GAG	CAA	GAC	AGA	GGC	AGA	TCC	GTG	CGA	2208
Glu	Thr	Glu	Gly	Gly	Val	Gly	Glu	Gln	Asp	Arg	Gly	Arg	Ser	Val	Arg	
		845					850					855				
TTA	GTG	AGC	GGA	TTC	TCA	GCT	CTT	GTC	TGG	GAG	GAC	CTC	CGG	AAC	CTG	2256
Leu	Val	Ser	Gly	Phe	Ser	Ala	Leu	Val	Trp	Glu	Asp	Leu	Arg	Asn	Leu	
	860					865					870					
TTG	ATC	TTC	CTC	TAC	CAC	CGC	TTG	ACA	GAC	TCA	CTC	TTG	ATA	CTG	AGG	2304
Leu	Ile	Phe	Leu	Tyr	His	Arg	Leu	Thr	Asp	Ser	Leu	Leu	Ile	Leu	Arg	
875					880					885				890		
AGG	ACT	CTG	GAA	CTC	CTG	GGA	CAG	AGT	CTC	AGC	AGG	GGA	CTG	CAA	CTA	2352
Arg	Thr	Leu	Glu	Leu	Leu	Gly	Gln	Ser	Leu	Ser	Arg	Gly	Leu	Gln	Leu	
				895				900						905		
CTG	AAT	GAA	CTC	AGA	ACA	CAC	TTG	TGG	GGA	ATA	CTT	GCA	TAT	TGG	GGA	2400
Leu	Asn	Glu	Leu	Arg	Thr	His	Leu	Trp	Gly	Ile	Leu	Ala	Tyr	Trp	Gly	
			910					915					920			
AAA	GAG	TTA	AGG	GAT	AGT	GCT	ATC	AGC	TTG	CTT	AAT	ACA	ACA	GCT	ATT	2448
Lys	Glu	Leu	Arg	Asp	Ser	Ala	Ile	Ser	Leu	Leu	Asn	Thr	Thr	Ala	Ile	
		925					930					935				
GTA	GTA	GCA	GAA	GGA	ACA	GAT	AGG	ATT	ATA	GAA	TTA	GCA	CAA	AGA	ATA	2496
Val	Val	Ala	Glu	Gly	Thr	Asp	Arg	Ile	Ile	Glu	Leu	Ala	Gln	Arg	Ile	
	940					945						950				
GGA	AGG	GGA	ATA	TTA	CAC	ATA	CCT	AGA	AGA	ATC	AGA	CAA	GGC	CTA	GAA	2544
Gly	Arg	Gly	Ile	Leu	His	Ile	Pro	Arg	Arg	Ile	Arg	Gln	Gly	Leu	Glu	
955					960					965				970		
AGA	GCA	CTG	ATA	TAA												2559
Arg	Ala	Leu	Ile													

(2) INFORMATION FOR SEQ ID NO: 18:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 852 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear



(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

Met Lys Val Met Gly Met Gln Ser Gly Trp Met Gly Met Lys Ser Gly
 1 5 10 15

Trp Leu Leu Phe Tyr Leu Leu Val Ser Leu Ile Lys Val Ile Gly Ser
 20 25 30

Glu Gln His Trp Val Thr Val Tyr Tyr Gly Val Pro Val Trp Arg Glu
 35 40 45

Ala Glu Thr Thr Leu Phe Cys Ala Ser Asp Ala Lys Ala His Ser Thr
 50 55 60

Glu Ala His Asn Ile Trp Ala Thr Gln Ala Cys Val Pro Thr Asp Pro
 65 70 75 80

Asn Pro Gln Glu Val Leu Leu Pro Asn Val Thr Glu Lys Phe Asn Met
 85 90 95

Trp Glu Asn Lys Met Ala Asp Gln Met Gln Glu Asp Ile Ile Ser Leu
 100 105 110

Trp Glu Gln Ser Leu Lys Pro Cys Val Lys Leu Thr Pro Leu Cys Val
 115 120 125

Thr Met Leu Cys Asn Asp Ser Tyr Gly Glu Glu Arg Asn Asn Thr Asn
 130 135 140

Met Thr Thr Arg Glu Pro Asp Ile Gly Tyr Lys Gln Met Lys Asn Cys
 145 150 155 160

Ser Phe Asn Ala Thr Thr Glu Leu Thr Asp Lys Lys Lys Gln Val Tyr
 165 170 175

Ser Leu Phe Tyr Val Glu Asp Val Val Pro Ile Asn Ala Tyr Asn Lys
 180 185 190

Thr Tyr Arg Leu Ile Asn Cys Asn Thr Thr Ala Val Thr Gln Ala Cys
 195 200 205

Pro Lys Thr Ser Phe Glu Pro Ile Pro Ile His Tyr Cys Ala Pro Pro
 210 215 220

Gly Phe Ala Ile Met Lys Cys Asn Glu Gly Asn Phe Ser Gly Asn Gly
 225 230 235 240

Ser Cys Thr Asn Val Ser Thr Val Gln Cys Thr His Gly Ile Lys Pro
 245 250 255

Val Ile Ser Thr Gln Leu Ile Leu Asn Gly Ser Leu Asn Thr Asp Gly
 260 265 270

Ile Val Ile Arg Asn Asp Ser His Ser Asn Leu Leu Val Gln Trp Asn



275					280					285					
Glu	Thr	Val	Pro	Ile	Asn	Cys	Thr	Arg	Pro	Gly	Asn	Asn	Thr	Gly	Gly
290						295					300				
Gln	Val	Gln	Ile	Gly	Pro	Ala	Met	Thr	Phe	Tyr	Asn	Ile	Glu	Lys	Ile
305					310					315					320
Val	Gly	Asp	Ile	Arg	Gln	Ala	Tyr	Cys	Asn	Val	Ser	Lys	Glu	Leu	Trp
				325					330					335	
Glu	Pro	Met	Trp	Asn	Arg	Thr	Arg	Glu	Glu	Ile	Lys	Lys	Ile	Leu	Gly
			340					345					350		
Lys	Asn	Asn	Ile	Thr	Phe	Arg	Ala	Arg	Glu	Arg	Asn	Glu	Gly	Asp	Leu
	355						360					365			
Glu	Val	Thr	His	Leu	Met	Phe	Asn	Cys	Arg	Gly	Glu	Phe	Phe	Tyr	Cys
370						375					380				
Asn	Thr	Ser	Lys	Leu	Phe	Asn	Glu	Glu	Leu	Leu	Asn	Glu	Thr	Gly	Glu
385					390					395					400
Pro	Ile	Thr	Leu	Pro	Cys	Arg	Ile	Arg	Gln	Ile	Val	Asn	Leu	Trp	Thr
				405					410					415	
Arg	Val	Gly	Lys	Gly	Ile	Tyr	Ala	Pro	Pro	Ile	Arg	Gly	Val	Leu	Asn
			420					425					430		
Cys	Thr	Ser	Asn	Ile	Thr	Gly	Leu	Val	Leu	Glu	Tyr	Ser	Gly	Gly	Pro
	435						440					445			
Asp	Thr	Lys	Glu	Thr	Ile	Val	Tyr	Pro	Ser	Gly	Gly	Asn	Met	Val	Asn
450						455					460				
Leu	Trp	Arg	Gln	Glu	Leu	Tyr	Lys	Tyr	Lys	Val	Val	Ser	Ile	Glu	Pro
465					470					475					480
Ile	Gly	Val	Ala	Pro	Gly	Lys	Ala	Lys	Arg	Arg	Thr	Val	Ser	Arg	Glu
				485					490					495	
Lys	Arg	Ala	Ala	Phe	Gly	Leu	Gly	Ala	Leu	Phe	Leu	Gly	Phe	Leu	Gly
		500						505					510		
Ala	Ala	Gly	Ser	Thr	Met	Gly	Ala	Ala	Ser	Ile	Thr	Leu	Thr	Val	Gln
		515					520						525		
Ala	Arg	Thr	Leu	Leu	Ser	Gly	Ile	Val	Gln	Gln	Gln	Asn	Ile	Leu	Leu
	530					535						540			
Arg	Ala	Ile	Glu	Ala	Gln	Gln	His	Leu	Leu	Gln	Leu	Ser	Ile	Trp	Gly
545					550					555					560
Ile	Lys	Gln	Leu	Gln	Ala	Lys	Val	Leu	Ala	Ile	Glu	Arg	Tyr	Leu	Arg
				565					570					575	



Asp Gln Gln Ile Leu Ser Leu Trp Gly Cys Ser Gly Lys Thr Ile Cys
 580 585 590
 Tyr Thr Thr Val Pro Trp Asn Glu Thr Trp Ser Asn Asn Thr Ser Tyr
 595 600 605
 Asp Thr Ile Trp Asn Asn Leu Thr Trp Gln Gln Trp Asp Glu Lys Val
 610 615 620
 Arg Asn Tyr Ser Gly Val Ile Phe Gly Leu Ile Glu Gln Ala Gln Glu
 625 630 635 640
 Gln Gln Asn Thr Asn Glu Lys Ser Leu Leu Glu Leu Asp Gln Trp Asp
 645 650 655
 Ser Leu Trp Ser Trp Phe Gly Ile Thr Lys Trp Leu Trp Tyr Ile Lys
 660 665 670
 Ile Ala Ile Met Ile Val Ala Gly Ile Val Gly Ile Arg Ile Ile Ser
 675 680 685
 Ile Val Ile Thr Ile Ile Ala Arg Val Arg Gln Gly Tyr Ser Pro Leu
 690 695 700
 Ser Leu Gln Thr Leu Ile Pro Thr Ala Arg Gly Pro Asp Arg Pro Glu
 705 710 715 720
 Glu Thr Glu Gly Gly Val Gly Glu Gln Asp Arg Gly Arg Ser Val Arg
 725 730 735
 Leu Val Ser Gly Phe Ser Ala Leu Val Trp Glu Asp Leu Arg Asn Leu
 740 745 750
 Leu Ile Phe Leu Tyr His Arg Leu Thr Asp Ser Leu Leu Ile Leu Arg
 755 760 765
 Arg Thr Leu Glu Leu Leu Gly Gln Ser Leu Ser Arg Gly Leu Gln Leu
 770 775 780
 Leu Asn Glu Leu Arg Thr His Leu Trp Gly Ile Leu Ala Tyr Trp Gly
 785 790 795 800
 Lys Glu Leu Arg Asp Ser Ala Ile Ser Leu Leu Asn Thr Thr Ala Ile
 805 810 815
 Val Val Ala Glu Gly Thr Asp Arg Ile Ile Glu Leu Ala Gln Arg Ile
 820 825 830
 Gly Arg Gly Ile Leu His Ile Pro Arg Arg Ile Arg Gln Gly Leu Glu
 835 840 845
 Arg Ala Leu Ile
 850



(2) INFORMATION FOR SEQ ID NO: 19:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGHT: 639 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: simple
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: ADN (genomic)

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..636

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

ATG GGA AAG ATT TGG TCA AAG AGC AGC CTA GTA GGA TGG CCA GAA ATC	48
Met Gly Lys Ile Trp Ser Lys Ser Ser Leu Val Gly Trp Pro Glu Ile	
855 860 865	
AGA GAA AGA ATG AGA AGA CAA ACG CAA GAA CCA GCA GTA GAG CCA GCA	96
Arg Glu Arg Met Arg Arg Gln Thr Gln Glu Pro Ala Val Glu Pro Ala	
870 875 880	
GTA GGA GCA GGA GCA GCT TCT CAA GAT CTA GCT AAT CGA GGG GCC ATC	144
Val Gly Ala Gly Ala Ala Ser Gln Asp Leu Ala Asn Arg Gly Ala Ile	
885 890 895 900	
ACC ATA AGA AAT ACT AGA GAC AAT AAT GAA AGT ATA GCT TGG CTA GAA	192
Thr Ile Arg Asn Thr Arg Asp Asn Asn Glu Ser Ile Ala Trp Leu Glu	
905 910 915	
GCA CAA GAA GAA GAA GAG GAA GTA GGC TTT CCA GTA CGC CCT CAG GTA	240
Ala Gln Glu Glu Glu Glu Glu Val Gly Phe Pro Val Arg Pro Gln Val	
920 925 930	
CCA TTA AGG CCA ATA ACC TAT AAA CAG GCT TTT GAT CTT TCC TTC TTT	288
Pro Leu Arg Pro Ile Thr Tyr Lys Gln Ala Phe Asp Leu Ser Phe Phe	
935 940 945	
TTA AAA GAT AAG GGG GGA CTG GAA GGG CTA GTT TGG TCC AGA AAA AGG	336
Leu Lys Asp Lys Gly Gly Leu Glu Gly Leu Val Trp Ser Arg Lys Arg	
950 955 960	
CAA GAT ATT CTA GAC CTC TGG ATG TAT CAC ACA CAA GGC ATC CTC CCT	384
Gln Asp Ile Leu Asp Leu Trp Met Tyr His Thr Gln Gly Ile Leu Pro	
965 970 975 980	
GAC TGG CAT AAC TAC ACA CCA GGG CCA GGA ATT AGA TAC CCC GTA ACC	432
Asp Trp His Asn Tyr Thr Pro Gly Pro Gly Ile Arg Tyr Pro Val Thr	
985 990 995	
TTT GGA TGG TGC TTC AAA CTA GTA CCA TTG TCA GCT GAA GAA GTA GAA	480



Phe Gly Trp Cys Phe Lys Leu Val Pro Leu Ser Ala Glu Glu Val Glu	
1000 1005 1010	
GAG GCT AAT GAA GGA GAC AAC AAT GCC CTC TTA CAC CCC ATA TGT CAA	528
Glu Ala Asn Glu Gly Asp Asn Asn Ala Leu Leu His Pro Ile Cys Gln	
1015 1020 1025	
CAT GGA GCA GAT GAT GAT CAT AAA GAA GTG TTG GTG TGG CGA TTT GAC	576
His Gly Ala Asp Asp Asp His Lys Glu Val Leu Val Trp Arg Phe Asp	
1030 1035 1040	
AGC TCC CTA GCA AGA AGA CAT GTA GCA AGA GAG CTG CAT CCG GAG TTT	624
Ser Ser Leu Ala Arg Arg His Val Ala Arg Glu Leu His Pro Glu Phe	
1045 1050 1055 1060	
TAC AAG AAC TGC TGA	639
Tyr Lys Asn Cys	

(2) INFORMATION FOR SEQ ID NO: 20:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 212 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

Met Gly Lys Ile Trp Ser Lys Ser Ser Leu Val Gly Trp Pro Glu Ile	
1 5 10 15	
Arg Glu Arg Met Arg Arg Gln Thr Gln Glu Pro Ala Val Glu Pro Ala	
20 25 30	
Val Gly Ala Gly Ala Ala Ser Gln Asp Leu Ala Asn Arg Gly Ala Ile	
35 40 45	
Thr Ile Arg Asn Thr Arg Asp Asn Asn Glu Ser Ile Ala Trp Leu Glu	
50 55 60	
Ala Gln Glu Glu Glu Glu Glu Val Gly Phe Pro Val Arg Pro Gln Val	
65 70 75 80	
Pro Leu Arg Pro Ile Thr Tyr Lys Gln Ala Phe Asp Leu Ser Phe Phe	
85 90 95	
Leu Lys Asp Lys Gly Gly Leu Glu Gly Leu Val Trp Ser Arg Lys Arg	
100 105 110	
Gln Asp Ile Leu Asp Leu Trp Met Tyr His Thr Gln Gly Ile Leu Pro	
115 120 125	
Asp Trp His Asn Tyr Thr Pro Gly Pro Gly Ile Arg Tyr Pro Val Thr	



130 135 140
 Phe Gly Trp Cys Phe Lys Leu Val Pro Leu Ser Ala Glu Glu Val Glu
 145 150 155 160
 Glu Ala Asn Glu Gly Asp Asn Asn Ala Leu Leu His Pro Ile Cys Gln
 165 170 175
 His Gly Ala Asp Asp Asp His Lys Glu Val Leu Val Trp Arg Phe Asp
 180 185 190
 Ser Ser Leu Ala Arg Arg His Val Ala Arg Glu Leu His Pro Glu Phe
 195 200 205
 Tyr Lys Asn Cys
 210

(2) INFORMATION FOR SEQ ID NO: 21:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGHT: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: simple
 - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: other nucleic acid
- (A) DESCRIPTION: /desc = «PRIMER»

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

ATTGCGTACT CACACTTCCG

20

(2) INFORMATION FOR SEQ ID NO: 22:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGHT: 17 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: simple
 - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: other nucleic acid
- (A) DESCRIPTION: /desc = «PRIMER»

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

GGCAAGCAGG GAGCTGG

17

(2) INFORMATION FOR SEQ ID NO: 23:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGHT: 18 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: simple



(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = «PRIMER»

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23:

TCCTTGAGCA GTCTGGAC

18

(2) INFORMATION FOR SEQ ID NO: 24:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGHT: 18 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: simple

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = «PRIMER»

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 24:

GAACAGGAGG ATTAGCAG

18

(2) INFORMATION FOR SEQ ID NO: 25:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGHT: 18 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: simple

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = «PRIMER»

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 25:

AGCAGAGGCT ATGTCACA

18

(2) INFORMATION FOR SEQ ID NO: 26:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGHT: 19 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: simple

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = «PRIMER»

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 26:



TGTAAGGCCCTAGAAAGAG

19

(2) INFORMATION FOR SEQ ID NO: 27:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGHT: 18 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: simple
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

- (A) DESCRIPTION: /desc = «PRIMER»

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 27:

ACAGAGAACT CTCTGTAC

18

(2) INFORMATION FOR SEQ ID NO: 28:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGHT: 18 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: simple
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

- (A) DESCRIPTION: /desc = «PRIMER»

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 28:

AAGAAAAGCA GTTGGTAC

18

(2) INFORMATION FOR SEQ ID NO: 29:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGHT: 17 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: simple
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

- (A) DESCRIPTION: /desc = «PRIMER»

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 29:

TTTCTTCCCT GTATGTC

17

(2) INFORMATION FOR SEQ ID NO: 30:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGHT: 18 base pairs
- (B) TYPE: nucleic acid



- (C) STRANDEDNESS: simple
- (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: other nucleic acid
 - (A) DESCRIPTION: /desc = «PRIMER»

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 30:

GTTATATGGA TTCTCAGG

18

- (2) INFORMATION FOR SEQ ID NO: 31:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGHT: 19 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: simple
 - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: other nucleic acid
 - (A) DESCRIPTION: /desc = «PRIMER»

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 31:

TGGCAGCACA TTATACTGG

19

- (2) INFORMATION FOR SEQ ID NO: 32:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGHT: 23 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: simple
 - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: other nucleic acid
 - (A) DESCRIPTION: /desc = «PRIMER»

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 32:

ATCATTTACC AGTACATGGA CGA

23

- (2) INFORMATION FOR SEQ ID NO: 33:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGHT: 18 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: simple
 - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: other nucleic acid
 - (A) DESCRIPTION: /desc = «PRIMER»

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 33:

TGTCAGGGGT CGTAAAGC

18



(2) INFORMATION FOR SEQ ID NO: 34:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGHT: 18 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: simple
 - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: other nucleic acid
 - (A) DESCRIPTION: /desc = «PRIMER»

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 34:

TCCTCTGGAT GGGATATG

18

(2) INFORMATION FOR SEQ ID NO: 35:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGHT: 18 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: simple
 - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: other nucleic acid
 - (A) DESCRIPTION: /desc = «PRIMER»

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 35:

TCTATCCAGG AATCAGAG

18

(2) INFORMATION FOR SEQ ID NO: 36:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGHT: 18 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: simple
 - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: other nucleic acid
 - (A) DESCRIPTION: /desc = «PRIMER»

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 36:

AATGAGATCT GCCCATAC

18

(2) INFORMATION FOR SEQ ID NO: 37:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGHT: 18 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: simple



(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = «PRIMER»

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 37:

TGACAGATAG GGGAAGAC

18

(2) INFORMATION FOR SEQ ID NO: 38:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGHT: 18 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: simple

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = «PRIMER»

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 38:

AACCGCCATT TGCACTGC

18

(2) INFORMATION FOR SEQ ID NO: 39:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGHT: 18 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: simple

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = «PRIMER»

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 39:

ACATGGACCG CCACAAGG

18

(2) INFORMATION FOR SEQ ID NO: 40:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGHT: 18 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: simple

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = «PRIMER»

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 40:

AGCAACAGAC ATACAGAC

18



(2) INFORMATION FOR SEQ ID NO: 41:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGHT: 18 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: simple
 (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: other nucleic acid
 (A) DESCRIPTION: /desc = «PRIMER»

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 41:

AAAGTAGTCC CACGTAGG

18

(2) INFORMATION FOR SEQ ID NO: 42:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGHT: 18 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: simple
 (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: other nucleic acid
 (A) DESCRIPTION: /desc = «PRIMER»

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 42:

ATATCCCAGT AGGTCAGG

18

(2) INFORMATION FOR SEQ ID NO: 43:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGHT: 18 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: simple
 (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: other nucleic acid
 (A) DESCRIPTION: /desc = «PRIMER»

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 43:

TCTAGCACTA ACAGCCTG

18

(2) INFORMATION FOR SEQ ID NO: 44:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGHT: 18 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: simple
 (D) TOPOLOGY: linear



- (ii) MOLECULE TYPE: other nucleic acid
 - (A) DESCRIPTION: /desc = «PRIMER»

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 44:

ACTCTTACTG CTCTGAGG

18

(2) INFORMATION FOR SEQ ID NO: 45:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGHT: 18 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: simple
 - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: other nucleic acid
 - (A) DESCRIPTION: /desc = «PRIMER»

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 45:

CCATAGTACA CTGTTACC

18

(2) INFORMATION FOR SEQ ID NO: 46:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGHT: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: simple
 - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: other nucleic acid
 - (A) DESCRIPTION: /desc = «PRIMER»

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 46:

CATAGCTATC GTTACAAAGC

20

(2) INFORMATION FOR SEQ ID NO: 47:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGHT: 18 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: simple
 - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: other nucleic acid
 - (A) DESCRIPTION: /desc = «PRIMER»

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 47:



TCATAATGGC AAAGCCTG

18

(2) INFORMATION FOR SEQ ID NO: 48:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGHT: 18 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: simple
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

- (A) DESCRIPTION: /desc = «PRIMER»

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 48:

CTATTCCACA TTGGTTCC

18

(2) INFORMATION FOR SEQ ID NO: 49:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGHT: 18 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: simple
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

- (A) DESCRIPTION: /desc = «PRIMER»

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 49:

ATTCTAGAAC CAGTCCAG

18

(2) INFORMATION FOR SEQ ID NO: 50:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGHT: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: simple
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

- (A) DESCRIPTION: /desc = «PRIMER»

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 50:

CCTTAGGGAT CAGCAAATCC

20

(2) INFORMATION FOR SEQ ID NO: 51:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGHT: 18 base pairs
- (B) TYPE: nucleic acid



- (C) STRANDEDNESS: simple
- (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: other nucleic acid
 - (A) DESCRIPTION: /desc = «PRIMER»

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 51:

TGGGACAGTC TGTGGAGC

18

- (2) INFORMATION FOR SEQ ID NO: 52:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGHT: 18 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: simple
 - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: other nucleic acid
 - (A) DESCRIPTION: /desc = «PRIMER»

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 52:

TTCTCAGCTC TTGTCTGG

18

- (2) INFORMATION FOR SEQ ID NO: 53:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGHT: 18 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: simple
 - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: other nucleic acid
 - (A) DESCRIPTION: /desc = «PRIMER»

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 53:

ATTAAGCAAG CTGATAGC

18

- (2) INFORMATION FOR SEQ ID NO: 54:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGHT: 16 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: simple
 - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: other nucleic acid
 - (A) DESCRIPTION: /desc = «PRIMER»

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 54:

TGTGCTTCTA GCCAAG

16



(2) INFORMATION FOR SEQ ID NO: 55:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGHT: 18 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: simple
 - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: other nucleic acid
 - (A) DESCRIPTION: /desc = «PRIMER»

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 55:

GCTCCATGTT GACATATG

18

(2) INFORMATION FOR SEQ ID NO: 56:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGHT: 18 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: simple
 - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: other nucleic acid
 - (A) DESCRIPTION: /desc = «PRIMER»

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 56:

AGAGAGACCC AGTACAAG

18

(2) INFORMATION FOR SEQ ID NO: 57:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGHT: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: simple
 - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: other nucleic acid
 - (A) DESCRIPTION: /desc = "PRIMER"

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 57:

ATAAAAGCAG CCGCTTCTCG

20

(2) INFORMATION FOR SEQ ID NO: 58:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGHT: 35 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: simple
 - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: peptide



(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 58:

Cys Thr Arg Pro Gly Asn Asn Thr Gly Gly Gln Val Gln Ile Gly Pro
 1 5 10 15

Ala Met Thr Phe Tyr Asn Ile Glu Lys Ile Val Gly Asp Ile Arg Gln
 20 25 30

Ala Tyr Cys
 35

(2) INFORMATION FOR SEQ ID NO: 59:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGHT: 35 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: simple
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 59:

Cys His Arg Pro Gly Asn Asn Thr Arg Gly Glu Val Gln Ile Gly Pro
 1 5 10 15

Gly Met Thr Phe Tyr Asn Ile Glu Asn Val Tyr Gly Asp Thr Arg Ser
 20 25 30

Ala Tyr Cys
 35

(2) INFORMATION FOR SEQ ID NO: 60:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGHT: 35 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: simple
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 60:

Cys Ile Arg Pro Gly Asn Arg Thr Tyr Arg Asn Leu Gln Ile Gly Pro
 1 5 10 15

Gly Met Thr Phe Tyr Asn Val Glu Ile Ala Thr Gly Asp Ile Arg Lys
 20 25 30

Ala Phe Cys
 35



(2) INFORMATION FOR SEQ ID NO: 61:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGHT: 35 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: simple
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 61:

Cys Thr Arg Pro Asn Asn Asn Thr Arg Lys Ser Val Arg Ile Gly Pro
1 5 10 15

Gly Gln Ala Phe Tyr Ala Thr Gly Asp Ile Ile Gly Asp Ile Arg Gln
 20 25 30

Ala His Cys
 35



MICROORGANISMS

Optional Sheet in connection with the microorganism referred to on page 3, line 30 of the description.

A. IDENTIFICATION OF DEPOSIT:

Further deposits are identified on an additional sheet ☐.

Name of depositary institution:

Collection Nationale de Cultures de Micro-organismes

Address of depositary institution (including postal code and country):

28 rue du Docteur Roux, 75724 PARIS CEDEX 15

Date of deposit:

July 2, 1996

Accession Number:

I-1753

B. ADDITIONAL INDICATIONS: (leave blank if not applicable). This information is continued on a separate attached sheet ☐.

"With regard to the nominations in which a European patent is applied for, until the publication of the mention of the grant of the European patent or until the date on which the application shall be refused or withdrawn or shall be deemed to be withdrawn, a sample of the deposited microorganism shall be available only by the issue of a sample to an expert nominated by the requester. (Rule 28.4) of the EPC)".

C. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE: (If the indications are not for all designated States)

ALL PCT CONTRACTING STATES

D. SEPARATE FURNISHING OF INDICATIONS: (leave blank if not applicable)

The indications listed below will be submitted to the International Bureau later: (Specify the general nature of the indications e.g., "Accession Number of Deposit")

E. ☐ This sheet was received with the international application when filed (to be checked by the receiving Office)

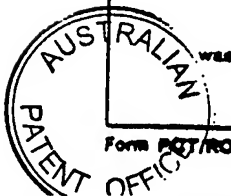
(illegible signature)

(Authorized Officer)

☐ The date of receipt (from the applicant) by the International Bureau is:

(Authorized Officer)

(January 1985)



THE CLAIMS DEFINING THE INVENTION ARE AS FOLLOWS:

1. An isolated non-M, non-O strain of HIV-1, a sample of which was deposited on 2 July 1996 under number
5 I-1753 (designated YBF30) in the Collection Nationale de Cultures de Microorganismes (National Collection of Microorganism Cultures) kept by the Pasteur Institute.
2. An isolated nucleic acid sequence, wherein the
10 sequence is derived from the strain according to Claim 1 and is selected from the group consisting of: SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19 and SEQ ID NOS:21 to 57, and wherein
15 said sequence is capable of hybridizing with a nucleic acid sequence which is derived from a non-M, non-O HIV-1 virus.
3. An isolated oligonucleotide wherein said
20 oligonucleotide comprises a nucleic acid sequence selected from the group consisting of SEQ ID NOS:21 to 57, and wherein said oligonucleotide is capable of being used as a primer or as a probe for detecting a non-M, non-O HIV-1 strain according to Claim 1.
- 25 4. An isolated non-M, non-O strain HIV-1 virus, wherein the virus exhibits the following characteristics:
 - (a) little or no serological reactivity with regard to proteins of the M and O groups and strong
30 serological reactivity with regard to proteins which are derived from the YBF30 strain according to Claim 1 or the CPZGAB SIV strain;
 - (b) absence of genomic amplification when using



H:\janc1\Keep\Speci\53272-98.doc 16/07/01

primers form the env and gag regions of the HIV-1 viruses of the M and O groups;

(c) genomic amplification in the presence of the primers which are derived from the YBF30 strain according to Claim 3; and

(d) greater than 70% sequence homology with the polynucleotide or polypeptide products of the envelope gene with regard to the corresponding polypeptide or polynucleotide products of the envelope gene of YBF30 strain.

5. An oligonucleotide which comprises a nucleic acid sequence selected from the group consisting of: SEQ ID NOS: 21 to 57, and wherein said oligonucleotide is capable of being used as a primer or as a probe for detecting a non-M, non-O strain of HIV-1 according to Claim 4.

6. A method of *in vitro* diagnosis of non-M, non-O strain HIV-1 virus comprising the steps of:

(a) providing a biological sample suspected of comprising a nucleic acid sequence to be detected;

(b) hybridizing the nucleic acid of (a) with at least one nucleic acid sequence according to Claim 2 or Claim 3; and

(c) detecting the presence of the hybridized nucleic acid sequence.

7. An isolated peptide capable of being expressed by a non-M, non-O strain of HIV-1 virus according to Claim 1 or Claim 4 or encoded by a nucleotide sequence according to Claim 2, wherein said peptide is capable of:

(a) being recognized by antibodies which are induced by a non-M, non-O HIV-1 virus according to Claim 1 or Claim 4, or a variant of this virus, and which are



present in a biological sample which is obtained following an infection with a non-M, non-O HIV-1 strain; and/or

(b) inducing the production of anti-non-M, non-O HIV-1 antibodies.

5

8. A peptide according to Claim 7, wherein said peptide is expressed by a nucleic acid sequence selected from the group consisting of: SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:58 and SEQ ID NO:20.

10

9. An immunogenic composition comprising one or more translation products of the nucleotide sequences according to Claim 2 and/or one of the peptides according to Claim 7 or Claim 8.

15

10. An isolated antibody directed against one or more of the peptides according to Claim 7 or Claim 8.

11. A method for the *in vitro* diagnosis of a non-M, non-O strain of HIV-1 virus, comprising the steps of:

20

- (a) providing a biological sample to be tested;
- (b) combining the sample of (a) with an

antibody according to Claim 10 [which may possibly be combined with anti-CPZGAB SIV antibodies]; and

25

- (c) detecting the presence of antibody-antigen complexes.

12. A method according to Claim 11, wherein the antibody of step (b) is also combined with anti-CPZGAB SIV antibodies.

30

13. A reagent for diagnosing a non-M, non-O strain



HIV-1 virus, comprising a nucleic acid or peptide sequence according to any one of Claims 2, 3, 7 or 8.

14. A method for screening and typing a non-M, non-O strain HIV-1 virus, comprising the steps of:

- (a) bringing a nucleic acid sequence selected from the group consisting of SEQ ID NOS: 2, 3, 5, 7, 9, 11, 13, 15, 17, 19 and 21 to 57 into contact with the nucleic acid of the virus to be typed; and
- 10 (b) detecting the hybrid which is formed.

15. A kit for diagnosing a non-M, non-O strain HIV-1 virus, comprising a reagent according to Claim 13.

15 16. An isolated virus according to claim 1, substantially as herein described with reference to any one of the examples or figures.

17. An isolated virus according to claim 4, substantially as herein described with reference to any one of the examples or figures.

Dated this 16th day of July 2001

INSTITUT NATIONAL DE LA SANTE ET DE LA RECHERCHE MEDICALE

25 (INSERM) and ASSISTANCE PUBLIQUE-HOPITAUX DE PARIS and
INSTITUT PASTEUR

By their Patent Attorneys

GRIFFITH HACK

Fellows Institute of Patent and

30 Trade Mark Attorneys of Australia



YLG	<i>ltr</i>	A T T G C G T A C T C A C A C T T C C G
LPBS.1	<i>ltr</i>	G G C A A G C A G G G A G C T G G
GAG Y	<i>ltr</i>	T C C T T G A G C A G T C T G G A C
AS1.1		
GAG Y	<i>gag</i>	G A A C A G G A G G A T T A G C A G
AS1		
Gag 6	<i>gag</i>	A G C A G A G G C T A T G T C A C A
GAG Y S1	<i>gag</i>	T G T A A G G C C C C T A G A A G A G
GAG Y	<i>gag</i>	A C A G A G A A C T C T C T G T A C
S1.1		
GAG Y	<i>gag</i>	A A G A A A A G C A G T T G G T A C
S1.2		
YRT AS	<i>pol</i>	T T T C T T C C C T G T A T G T C
1.3		
YRT AS1.2	<i>pol</i>	G T T A T A T G G A T T C T C A G G
YRT AS1.1	<i>pol</i>	T G G C A G C A C A T T A T A C T G G
YRT2	<i>pol</i>	A T C A T T T A C C A G T A C A T G G A C G A
YRT AS1	<i>pol</i>	T G T C A G G G G T C G T A A A G C
YRT2-1	<i>pol</i>	T C C T C T G G A T G G G A T A T G
YRT2-2	<i>pol</i>	T C T A T C C A G G A A T C A G A G
YRT-3	<i>pol</i>	A A T G A G A T C T G C C C A T A C
YRT2-4	<i>pol</i>	T G A C A G A T A G G G G A A G A C
4481-1	<i>pol</i>	A A C C G C C A T T T G C A C T G C
4481-2	<i>pol</i>	A C A T G G A C C G C C A C A A G G
4235.1	<i>pol</i>	A G C A A C A G A C A T A C A G A C
4235.2	<i>vif</i>	A A A G T A G T C C C A C G T A G G
4235.3	<i>tat</i>	A T A T C C C A G T A G G T C A G G
4235.4	<i>tat</i>	T C T A G C A C T A A C A G C C T G
SK69.6	<i>env</i>	A C T C T T A C T G C T C T G A G G
SK69.5	<i>env</i>	C C A T A G T A C A C T G T T A C C
SK69.4	<i>env</i>	C A T A G C T A T C G T T A C A A A G C
SK69.3	<i>env</i>	T C A T A A T G G C A A A G C C T G
SK69.2	<i>env</i>	C T A T T C C A C A T T G G T T C C
SK69.1	<i>env</i>	A T T C T A G A A C C A G T C C A G
SK68.1	<i>env</i>	C C T T A G G G A T C A G C A A A T C C
SK68.2	<i>env</i>	T G G G A C A G T C T G T G G A G C
SK68.3	<i>env</i>	T T C T C A G C T C T T G T C T G G
LSI AS1.3	<i>nef</i>	A T T A A G C A A G C T G A T A G C
LSI AS1.2	<i>nef</i>	T G T G C T T C T A G C C A A G
LSI AS 1.1	<i>ltr</i>	G C T C C A T G T T G A C A T A T G
LSi A1	<i>ltr</i>	A G A G A G A C C C A G T A C A A G
YLP A	<i>ltr</i>	A T A A A A G C A G C C G C T T C T C G

FIGURE 1

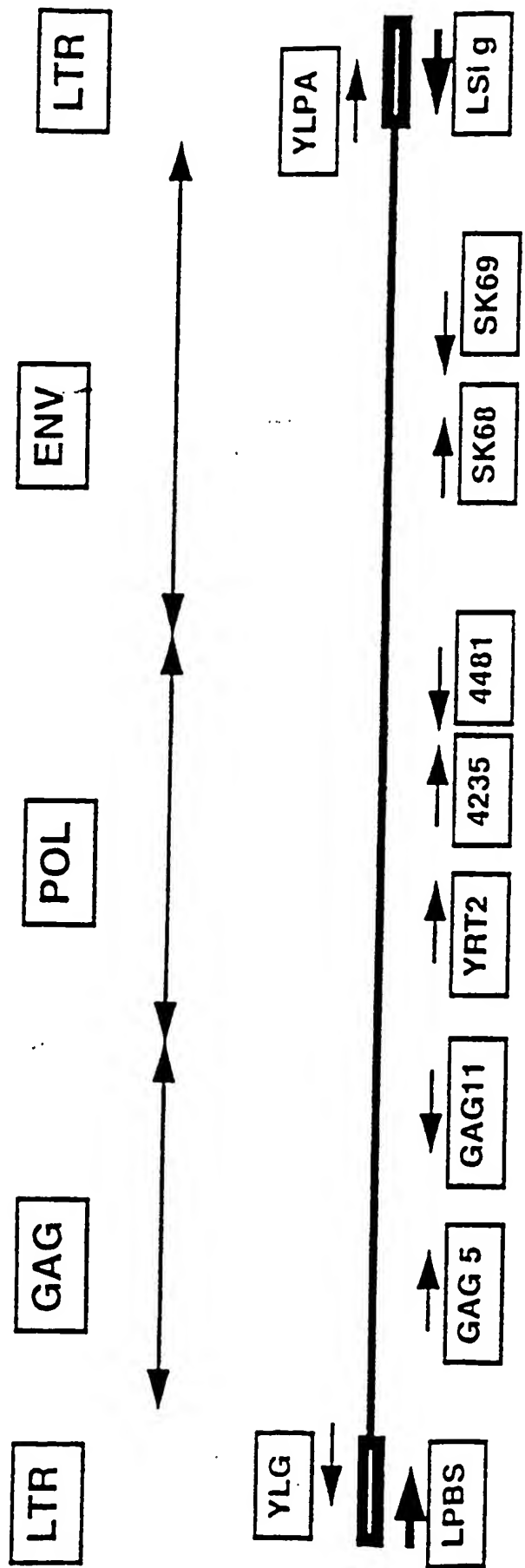
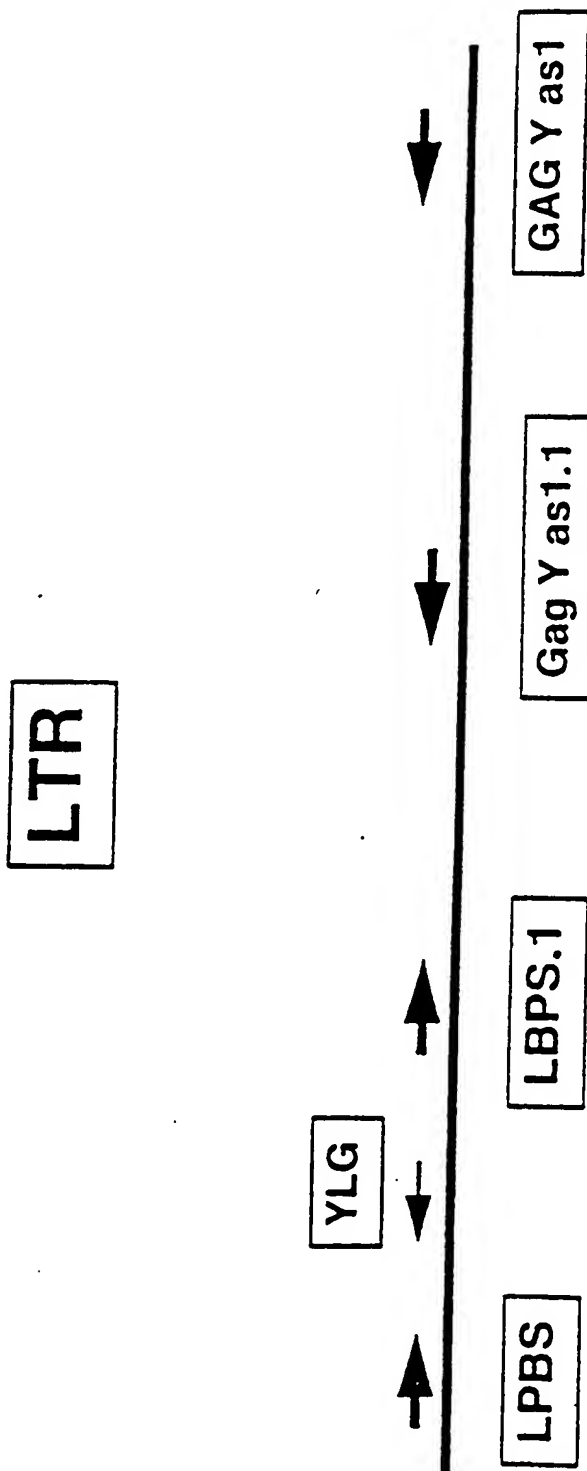


FIGURE 2

FIGURE 3

GAG

4/20

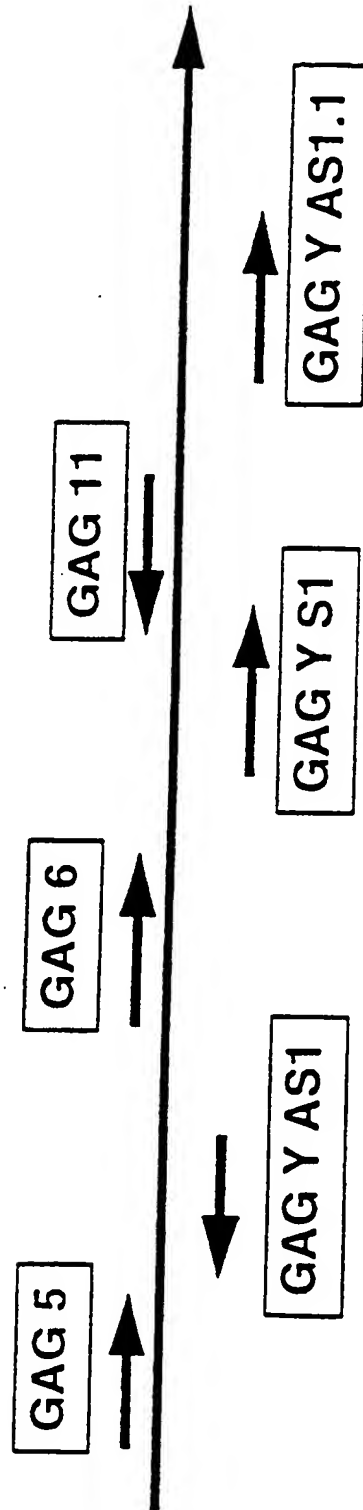


FIGURE 4

POL

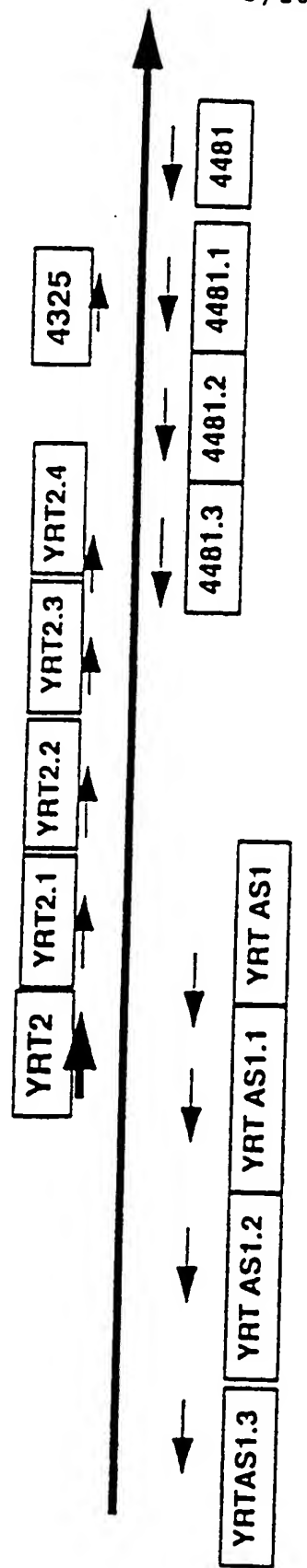


FIGURE 5

ENV

V3

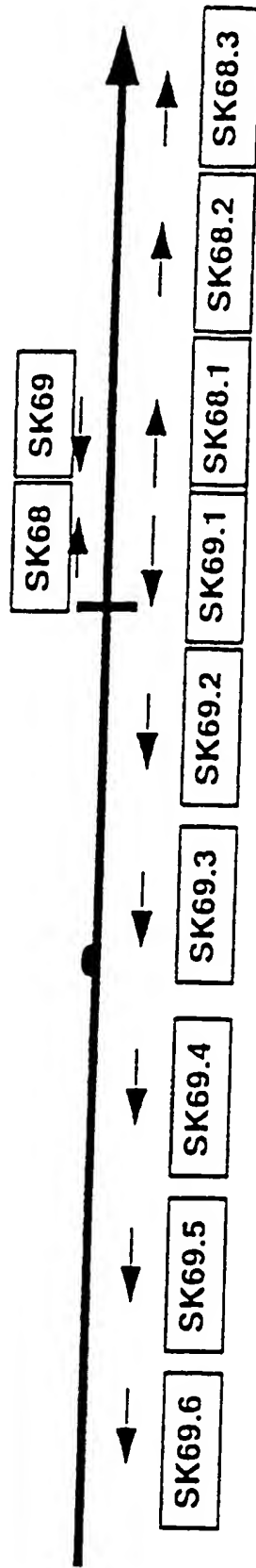


FIGURE 6

LTR

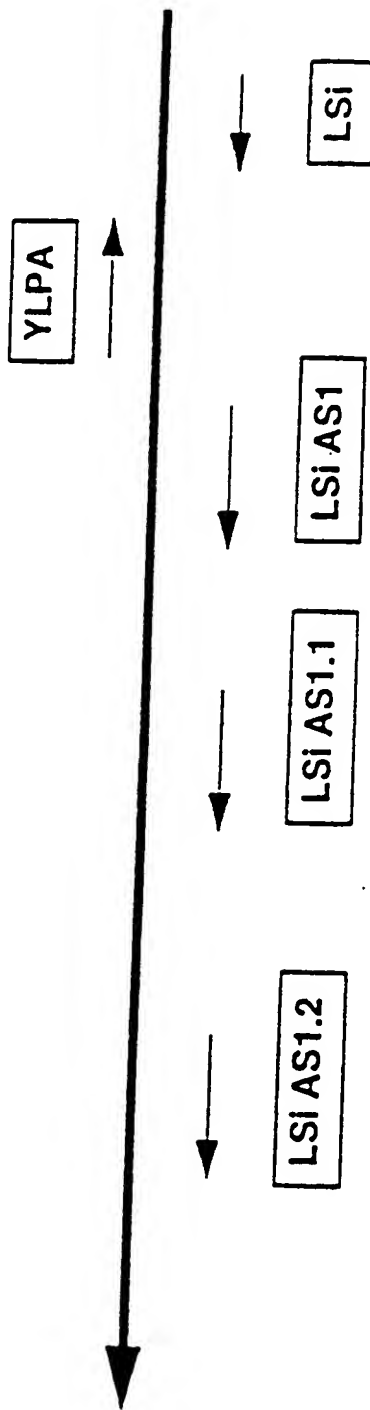


FIGURE 7

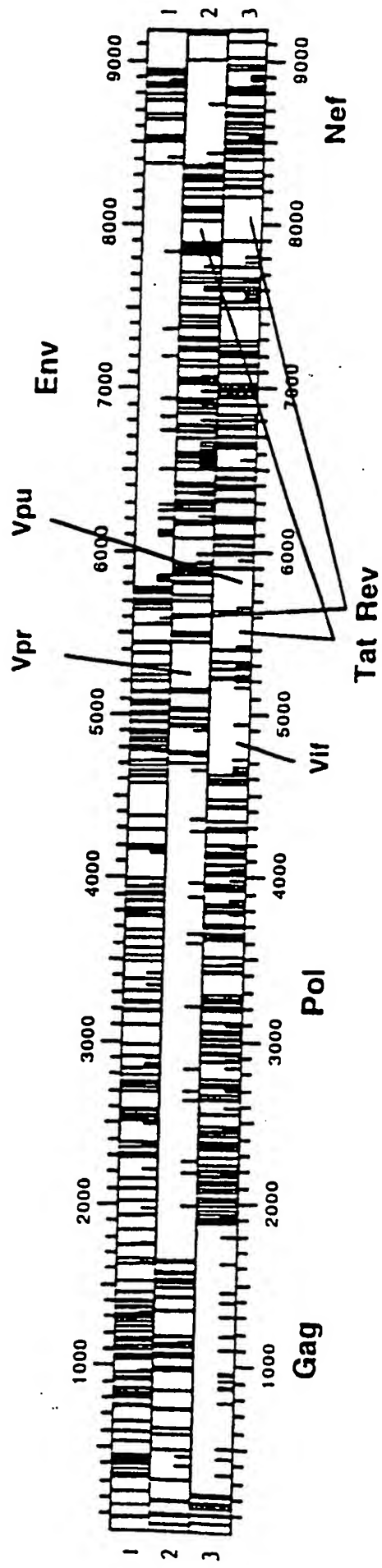


FIGURE 8

YBF30 LTR

464 nt after "bootstrapping"

PHYLIP n-j tree with "bootstrap" values (100 "bootstraps")

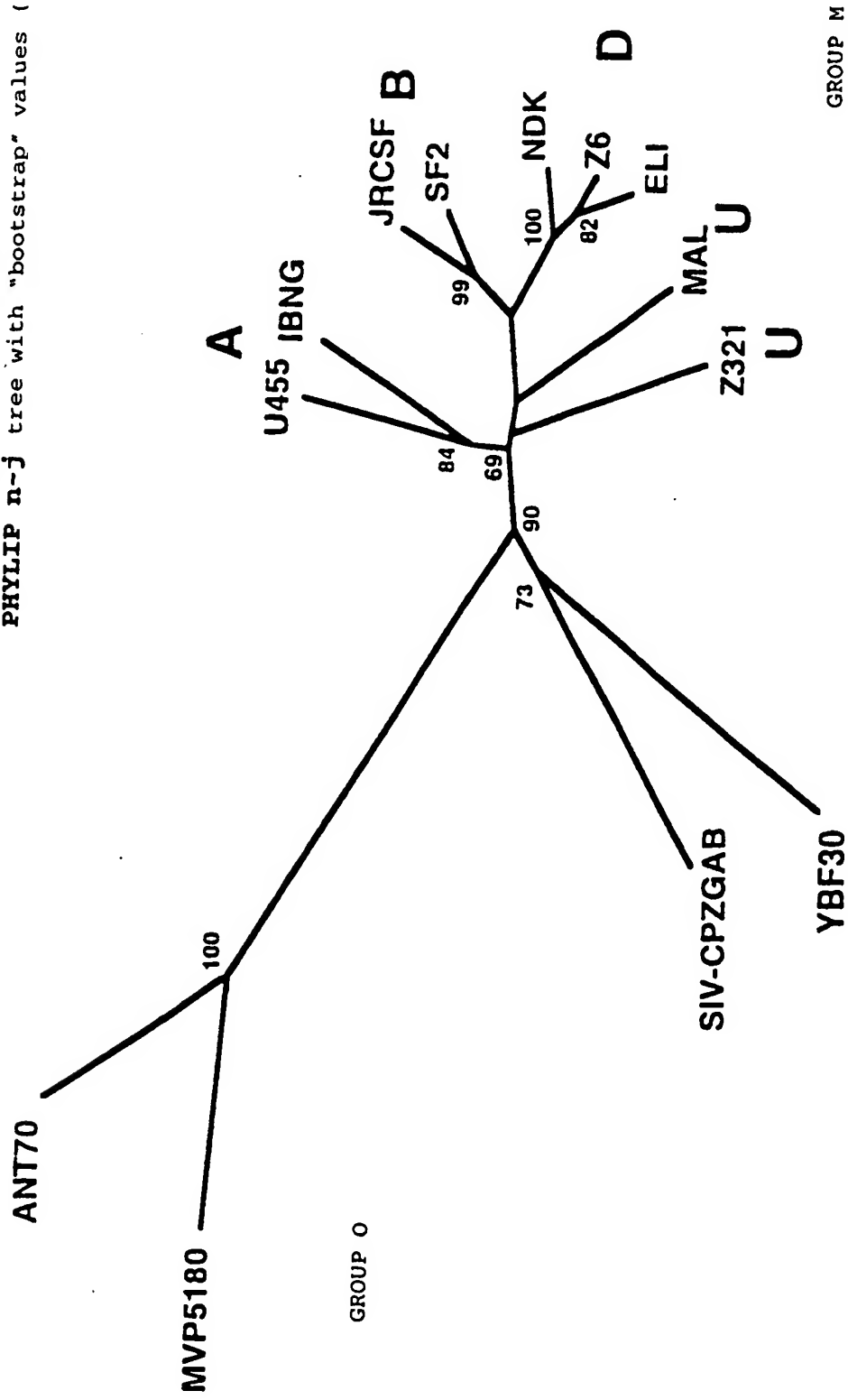
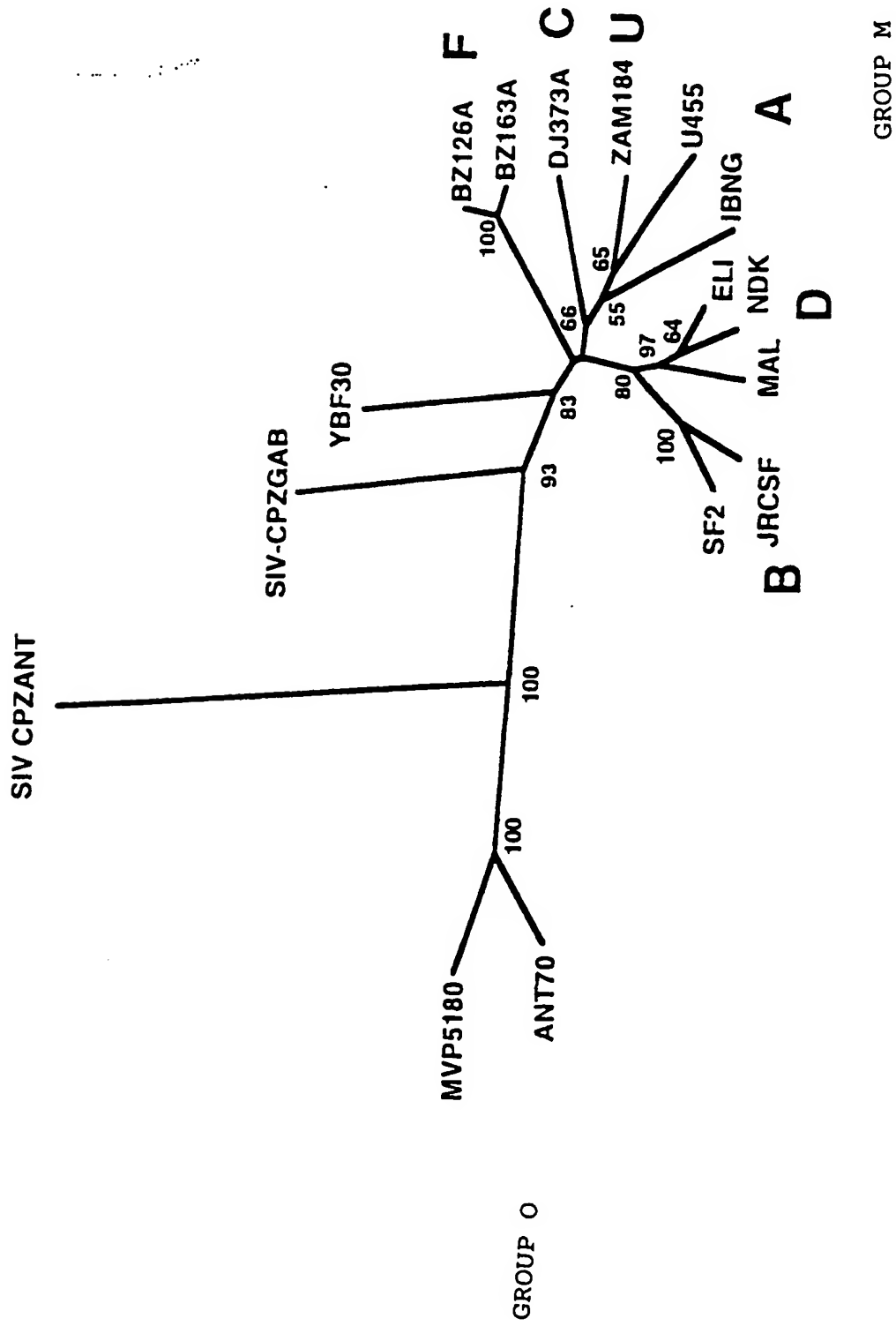


FIGURE 9



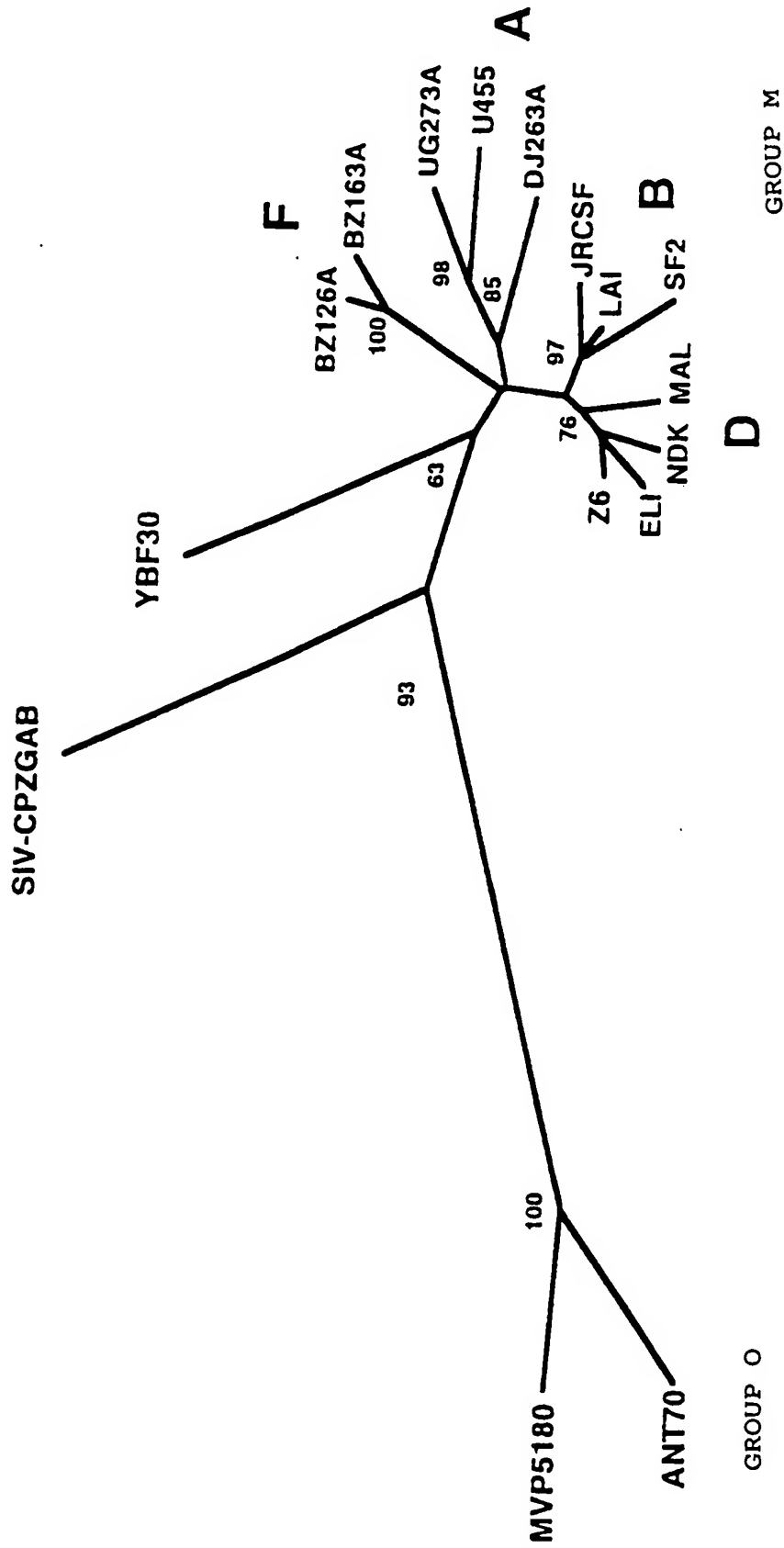


FIGURE 12

YBF30 Rev

296 nt after "gapstripping"

PHYLIP n-j tree with "bootstrap" values (100 "bootstraps")

YBF30 Vif

561 nt after "gapstripping"

PHYMLIP n-j tree with "bootstrap" values (100 "bootstraps")

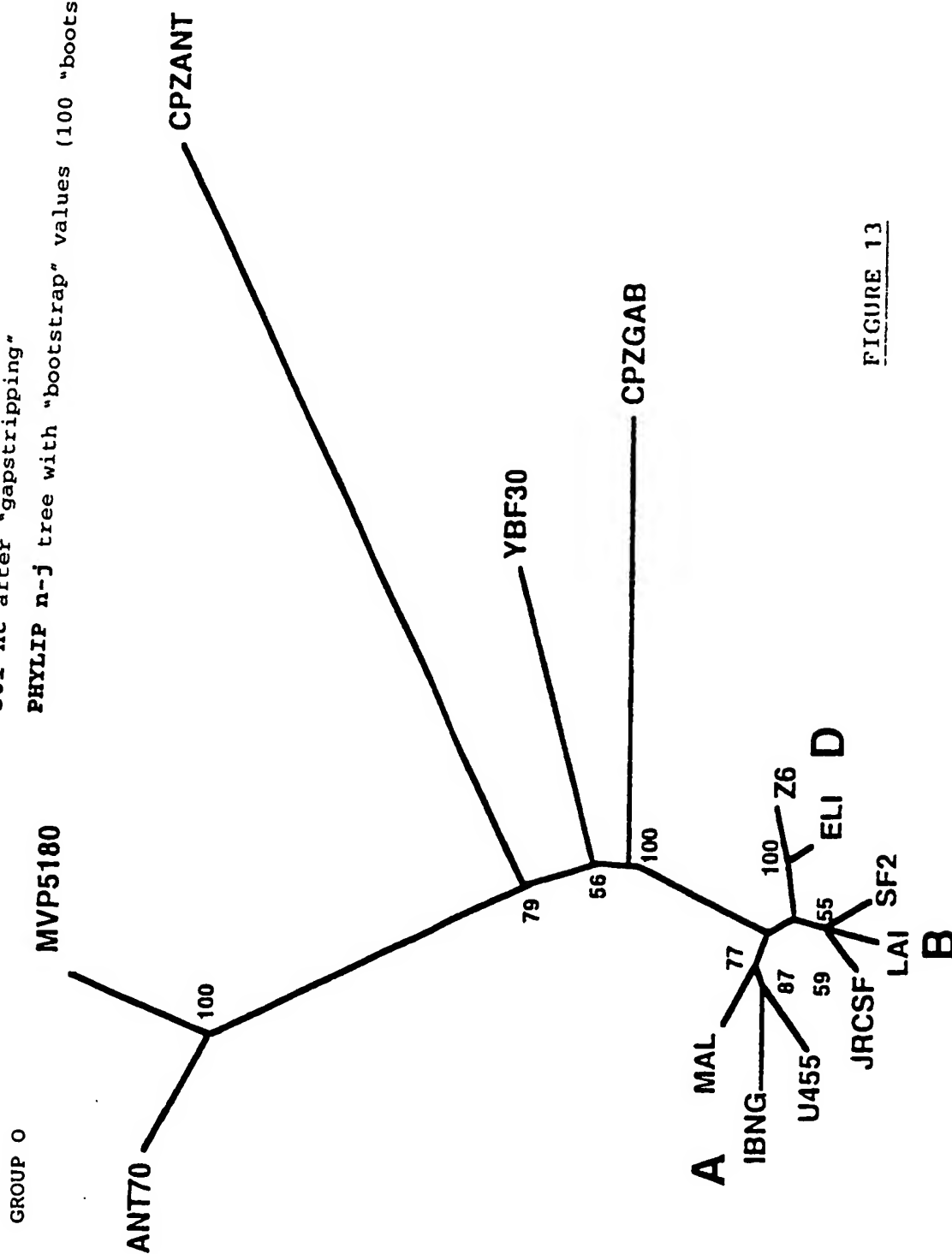
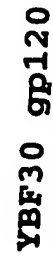


FIGURE 13

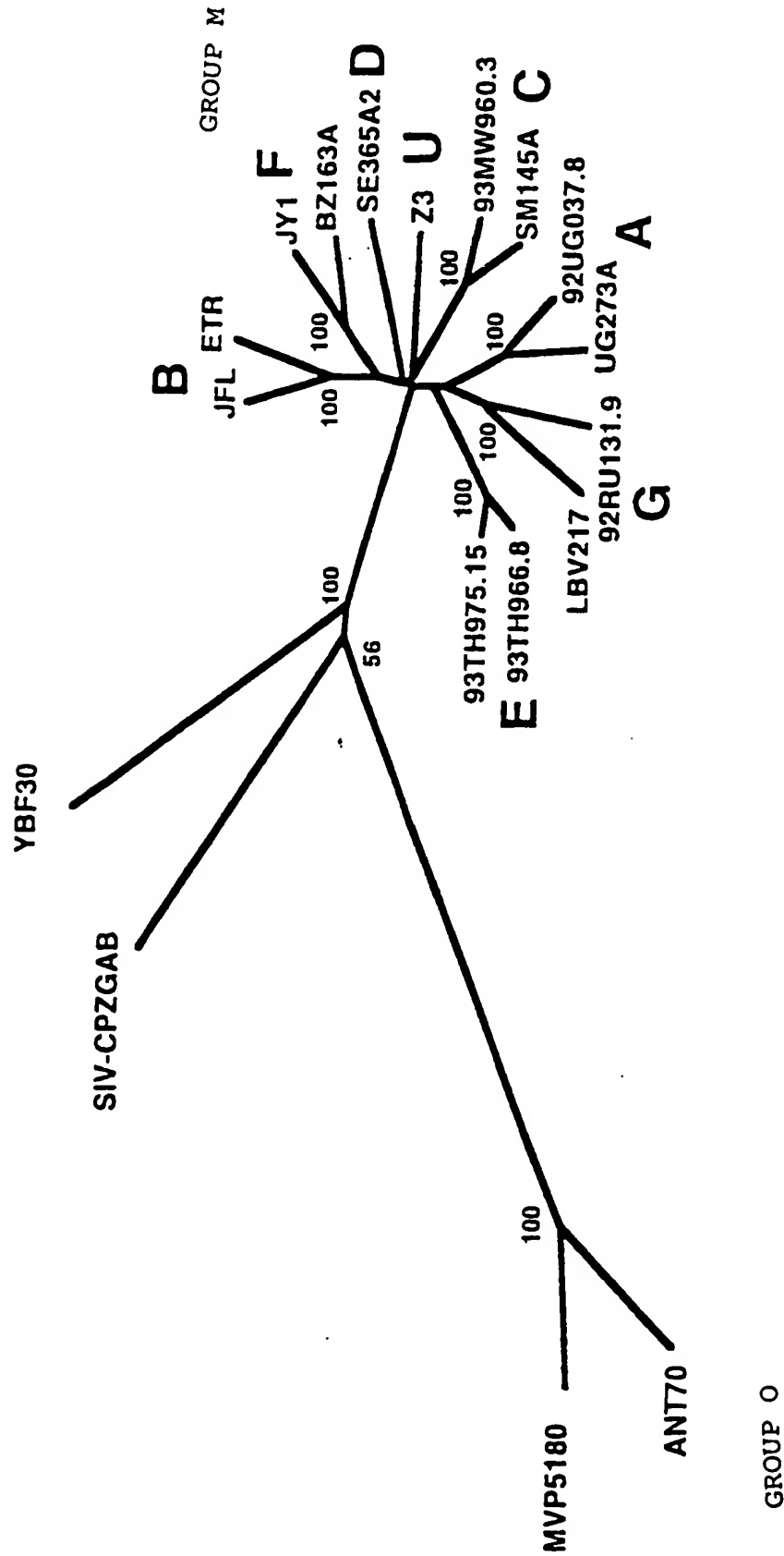


1317 nt after "gapstripping"

PHYLLIP n-j tree with "bootstrap" values (100 replicates)

Kimura distances, transition/transversion = 1.8

FIGURE 14



YBF30 gp41

988 nt after "gapstripping"

PHYLIP n-j tree with "bootstrap" values (100 "bootstraps"):

FIGURE 15

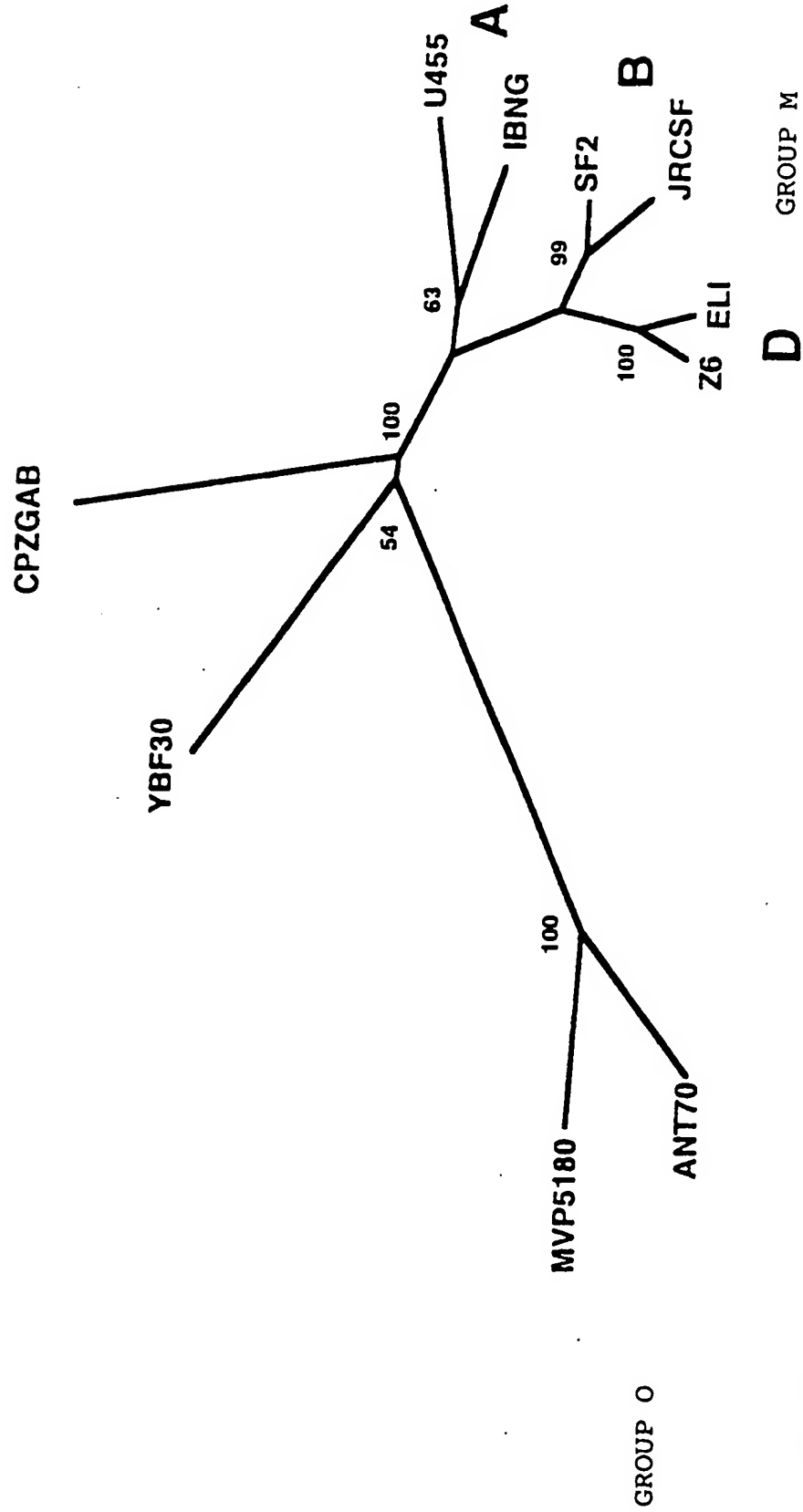
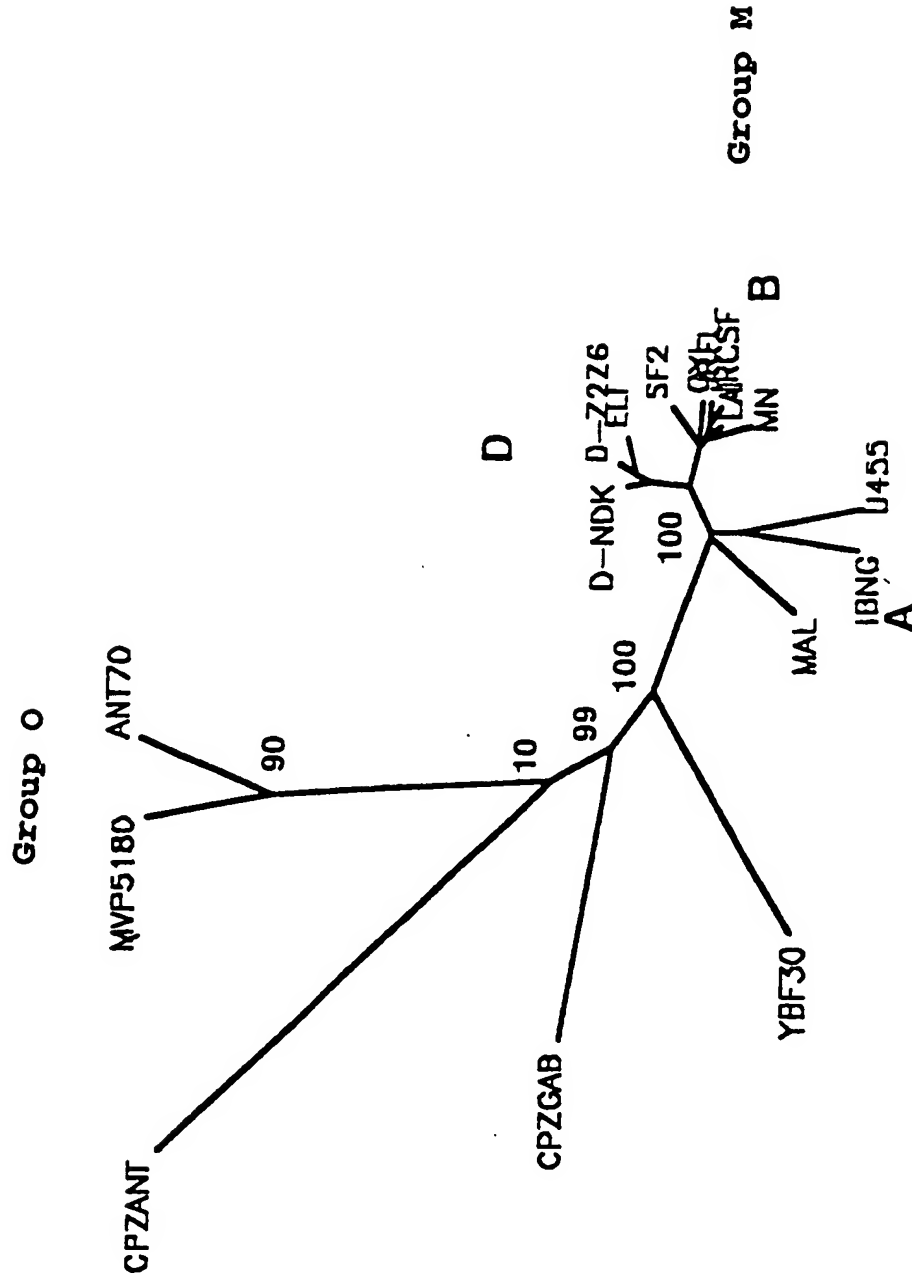


FIGURE 16

PHYLIP n-j tree with "bootstrap" values (100 "bootstraps")

YBF30 POL**Phylip Fitch 1867 nt after "gapstripping"**

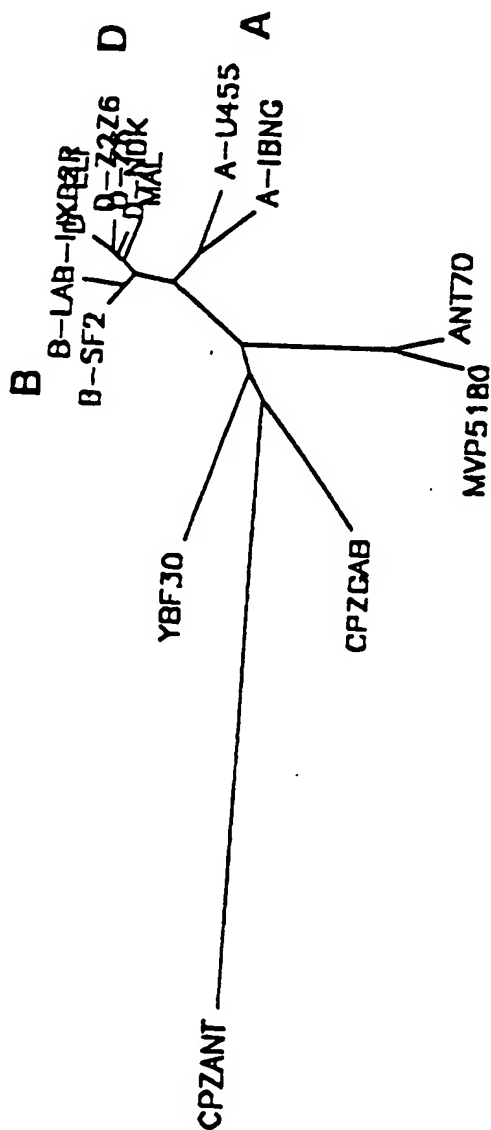
(100 "bootstraps")

**FIGURE 17**

YBF30 VPR

Phylip nj, 315 nt after "gapstripping"

Group M



Group O

FIGURE 18

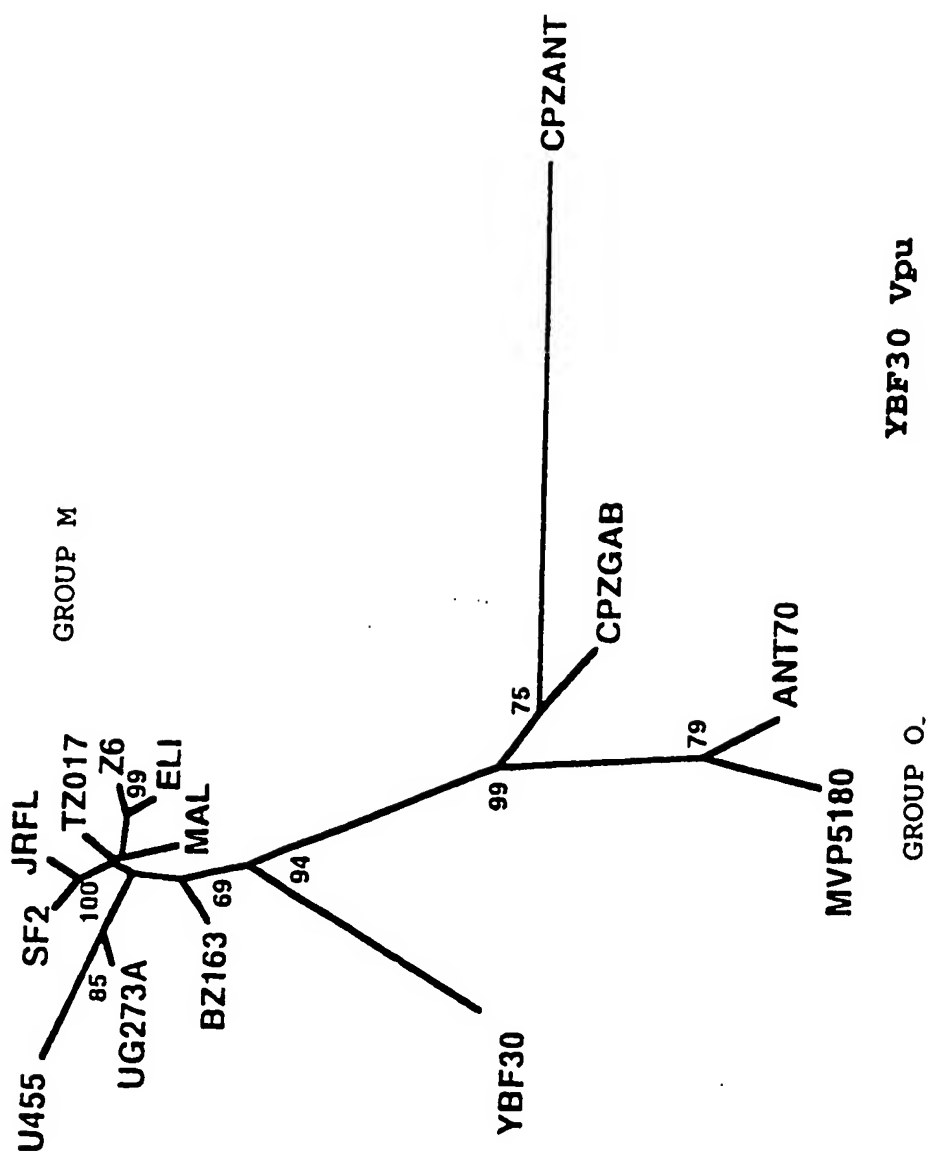


FIGURE 19

Percentage genetic distance between YBF30 and HIV-1/CPZSIV

	Gag	Pol	Vif	Vpr	Vpu	Tat	Rev	Env gp120	Nef
HIV-1 M	30-33	22-24	27.5-30	27-30	66.6-80	22-27.6	33.8-42	50-53	34.6-39
HIV-1 O	37-38	33-34	42-45.6	32-36	>100	46-47.7	80-88	73-74	52.8-53
CPZGAB	32	26.8	40.3	28.8	>100	27.8	56.8	50	33.7
CPZANT	45	41.2	57.1	57.4	>100	55	ND*	74.5	ND*

* ND: not determined

FIGURE 20

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☐ FADED TEXT OR DRAWING
- ☒ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.